ONTARIO

SEMINAR ON GEOCODING

ONTARIO STATISTICAL CENTRE
ECONOMIC AND STATISTICAL SERVICES DIVISION
DEPARTMENT OF TREASURY AND ECONOMICS

HON. W. DARCY McKEOUGH
TREASURER OF ONTARIO AND
MINISTER OF ECONOMICS

H. IAN MACDONALD
DEPUTY MINISTER

SEMINAR
ON
GEOCODING

Sponsored by

Ontario Statistical Centre

Economic and Statistical Services Division

Department of Treasury and Economics

Macdonald Block

Queen's Park

Toronto, Ontario

September 18, 1970

FOREWORD

The rapid growth and increasing industrialization of urban areas in recent decades have led slowly but surely to recognition of the fact that there is much need for more detailed urban information, particularly on a small area basis. Data which have been traditionally gathered by fixed administrative or political units are not capable of providing extensive information about changing urban environments.

Recent advances in data processing techniques have made possible methods to deal with this problem. Geocoding is one of these computer techniques which assigns geographical coordinates to property addresses, in the data input file, thereby allowing retrieval of information by user-specified areas.

The Ontario Statistical Centre has been conducting research and feasibility studies in the application of goeocoding to urban and rural areas. On September 18, 1970, a Seminar on Geocoding was held for the purpose of exposing geocoding concepts to government agencies and ensuring the development of a coordinated geocoding program

FOREWORD

The rapid growth and increasing industrialization of urban areas in recent decades have led slowly but surely to recognition of the fact that there is much need for more detailed urban information, particularly on a small area basis. Data which have been traditionally gathered by fixed administrative or political units are not capable of providing adequate information about changing urban environments.

Recent advances in data processing techniques have made possible methods to deal with this problem. Geocoding is one of these computer techniques which assigns geographical coordinates to property addresses, in the data input file, thereby allowing retrieval of information by user-specified areas.

The Ontario Statistical Centre has been conducting research and feasibility studies in the application of geocoding to urban and rural areas. On September 11, 1970, a Seminar on Geocoding was held for the purpose of acquainting interested persons in government agencies and exploring the development of a coordinated geocoding program

for the Ontario Government. The papers delivered at this Seminar illustrate the present state of the art of Geocoding.

We are deeply indebted to the speakers for their contribution in making this Seminar most instructive, and to the participants for their interest and cooperation in furthering the progress of this project. Finally, thanks should be extended to the Department of Highways for their assistance in organizing the Geocoding Seminar.

K. Cheng

K. Cheng
Director
Ontario Statistical Centre

ATTENDANCE

"Seminar on Geocoding"

September 18, 1970

St. Clair, Thames and Erie Room
2nd Floor, Macdonald Block, Queen's Park

Toronto, Ontario

Representatives from
The Ontario Government and Other Agencies

Mr. K. McCutcheon	Prime Minister's Office
Dr. W.C. Pfeiffer	Agriculture and Food
Mr. C.M. Riach	Agriculture and Food
Mr. S. Harris	Civil Service Commission
Mr. A. Rolavs	Education
Dr. J.A. Keddy	Education
Mr. M.E. Plewes	Energy and Resources
Mr. W.S. Preston	Ontario Hydro Commission
Mr. G. Scanlon	Water Resources Commission
Mr. F. Fleischer	Water Resources Commission
Mr. M. Holy	Water Resources Commission
Mrs. D. Stafl	Financial and Commercial Affairs
Dr. J. Smiley	Health
Mrs. H. Bain	Health
Mr. Harper	Health
Mrs. C. Rogos	Health
Mrs. B. Lang	Medical Services Insurance Division

Mr. K.Y. Shen	Highways
Mr. L. Schwabl	Highways
Mr. B. Wright	Highways
Mr. R.G. Hasenclever	Justice
Mr. L. Haywood	Labour
Mr. Horst Steibert	Labour
Mr. W. Genns	Lands and Forests
Mr. M. Berman	Lands and Forests
Mr. N.A. Vetere	Secretary and Citizenship
Mr. M.A. Brown	Secretary and Citizenship
Miss J.C. Downing	Municipal Affairs
Mr. M.H. Sinclair	Municipal Affairs
Mr. D. Montgomery	Municipal Affairs
Mr. D. Williams	Municipal Affairs
Mr. F. Watty	Municipal Affairs
Mr. B. Moody	Public Works
Mr. W.J. Smithson	Revenue
Mr. D. O'Connor	Revenue
Mr. K. Nash	Social and Family Services
Mr. T. Spearin	Tourism and Information
Mr. G. Angst	Tourism and Information
Mr. B. Tangney	Trade and Development
Mr. D. Dickinson	Trade and Development
Mrs. S. Carey	Ontario Housing Corporation
Mrs. I. Kostir	Ontario Housing Corporation

Mr. N. Choate	Ontario Housing Corporation
Mr. W.A. Furlonger	Transport
Mr. D.C. MacKinnon	Treasury and Economics
Mr. D.I. Stevenson	Treasury and Economics
Mr. B. Cahoon	Treasury and Economics
Mr. P. Cook	Treasury and Economics
Mr. C.L. Fan	Treasury and Economics
Mr. N. Ferik	Treasury and Economics
Miss D. Fetterly	Treasury and Economics
Miss B. Joyner	Treasury and Economics
Mr. M. Kogler	Treasury and Economics
Mrs. M.B. Levitt	Treasury and Economics
Mr. D. MacInnes	Treasury and Economics
Mr. A. MacKinnon	Treasury and Economics
Mr. G. McAllister	Treasury and Economics
Mr. H. McGonigal	Treasury and Economics
Mr. P. Mattson	Treasury and Economics
Mr. A. Noseworthy	Treasury and Economics
Mrs. M. Roderigues	Treasury and Economics
Mrs. H. Salisbury	Treasury and Economics
Mr. S.N. Sharma	Treasury and Economics
Miss L. Simpson	Treasury and Economics
Dr. J. Singh	Treasury and Economics
Mr. G.Z. Szabo	Treasury and Economics
Mr. D. Tanas	Treasury and Economics
Mr. S. Wright	Treasury and Economics

Mr. R. Grant	Treasury Board Secretariat
Mr. W. Nuss	Treasury Board Secretariat
Mrs. T. Anderson	University Affairs
Mr. Bellis	City of Toronto
Mr. D. Hillman	City of Toronto
Mr. B. Bailey	City of Hamilton
Mr. M.A. Khan	City of Hamilton
Mr. L. Lansa	City of Hamilton
Mr. J.W. VanLoon	City of Hamilton
Mr. R. Ion	Dominion Bureau of Statistics
Mr. W. Podehl	Dominion Bureau of Statistics
Professor E. Ferentzy	University of Toronto

Table of Contents

Page

Mr. H.I. Macdonald
Deputy Minister
Ontario Department of
Treasury and Economics

OPENING ADDRESS

The attendance at our first inter-departmental session to discuss the potential application of Geocoding in Ontario is indeed encouraging. As you will note from the program that has been distributed to you today, we have a lengthy agenda which includes a panel discussion and several guest speakers who are familiar with Geocoding and its applications. Perhaps I should begin the proceedings this morning by introducing myself. My name is Ian Macdonald of the Department of Treasury and Economics.

Government agencies are, as you know, presently involved in the collection of an ever-increasing amount of statistical data for research and planning. I am sure that we can anticipate that this trend will grow along with the tendency to greater urbanization in Canada, creating in turn a demand for more data to facilitiate urban and regional planning. For example, the Economic Council of Canada has estimated that over 80 percent of the 25 million population forecast for Canada in 1980 will reside in urban areas. Whether that is a desirable thing or not, and whether that will actually come about, will depend upon Government policies as much as anything else. Those of you who are familiar with

the ingredients of the Government's Regional Development Program know the effort which is being exerted to propose a shift in some of those trends in an attempt to bring about an ever-increasing decentralization of people in the Province. Whatever the outcome of these movements and these policies, I am sure we can anticipate, among ourselves at least, the needs for more and better data.

Traditionally, and until quite recently, data collection methods consisted of the assignment of numeric codes to basic geographic units, such as counties, municipalities, electoral districts, and enumeration areas. Based on the traditional coding concept, it is virtually impossible to compile or retrieve data for irregular geographic areas or for very small areas within traditional geographical boundaries. However, harnessing the recent advances in computer technology, it is now possible to develop and apply much more sophisticated methods with the result that techniques such as Geocoding have become prominent, or are about to become prominent in our Governmental application.

The purpose of this seminar is to discuss the various aspects of Geocoding, and to discover how it may assist you in your research and planning activities. I would like to stress, in particular, that the reason for arranging the seminar in this manner, and at this time, is that we not

plunge ahead with geocoding on our own without applying two stringent tests. Firstly, testing whether it will be of use to you, and secondly, attempting to discover to what extent you wish to participate in it.

Today you will hear from Mr. John Weldon, Chief of General Survey Systems, Dominion Bureau of Statistics, about the Geocoding System developed by his group for the 1971 population and household census. These developments are being closely followed by our central agency, the Ontario Statistical Centre. Also, on today's agenda, other speakers are broadly representative of their respective realms: Dr. Richard Thoman, Director, Regional Development Branch of the Department of Treasury and Economics, Professor E.M. Horwood, Professor of Civil Engineering of the University of Washington, whom we particularly welcome today, Professor R. McDaniel of the Department of Geography, University of Western Ontario, Mr. D.C. Symons, Chief of Computer Services for the National Capital Commission in Ottawa and Mr. David Weeks, Senior Programmer with the Department of Highways.

In advance, may I extend my own thanks to each one for his participation, and to each of you for agreeing to spend this day with us.

Dr. R.S. Thoman
Director
Regional Development Branch
Department of Treasury and Economics

'GENERAL BACKGROUND'
(Posing the problems of
requirements for data on small area basis)

Mr. Macdonald, Mr. Schnick, Ladies and Gentlemen, I trust you will pardon me this morning if we speak on some problems which are not quite as indicated on the Agenda, but which are certainly problems of Geocoding from the personal experience that we have had in Canada, and particularly in the Regional Development Branch. I think, perhaps, they can be stated more meaningfully if we approach them in a more or less applied fashion.

First of all, previous to working for the province, in a study which Mr. Maurice Yates and I produced at Queen's, we were asked to delimit the Georgian Bay Region. This was the old problem of designated areas, those who are inside a designated area are happy, those who are outside are not happy, and there is always the problem of how the boundary is to be drawn. We found, among other things, the need to look to Geocoding. We said in the report that there should be data for small areas, and optimally provided on a monthly basis. The data should refer to places of residence and not places of employment. The spatial size reporting units should range as follows:-

Census information and sample data should be available on block faces for all urban and sub-urban areas.

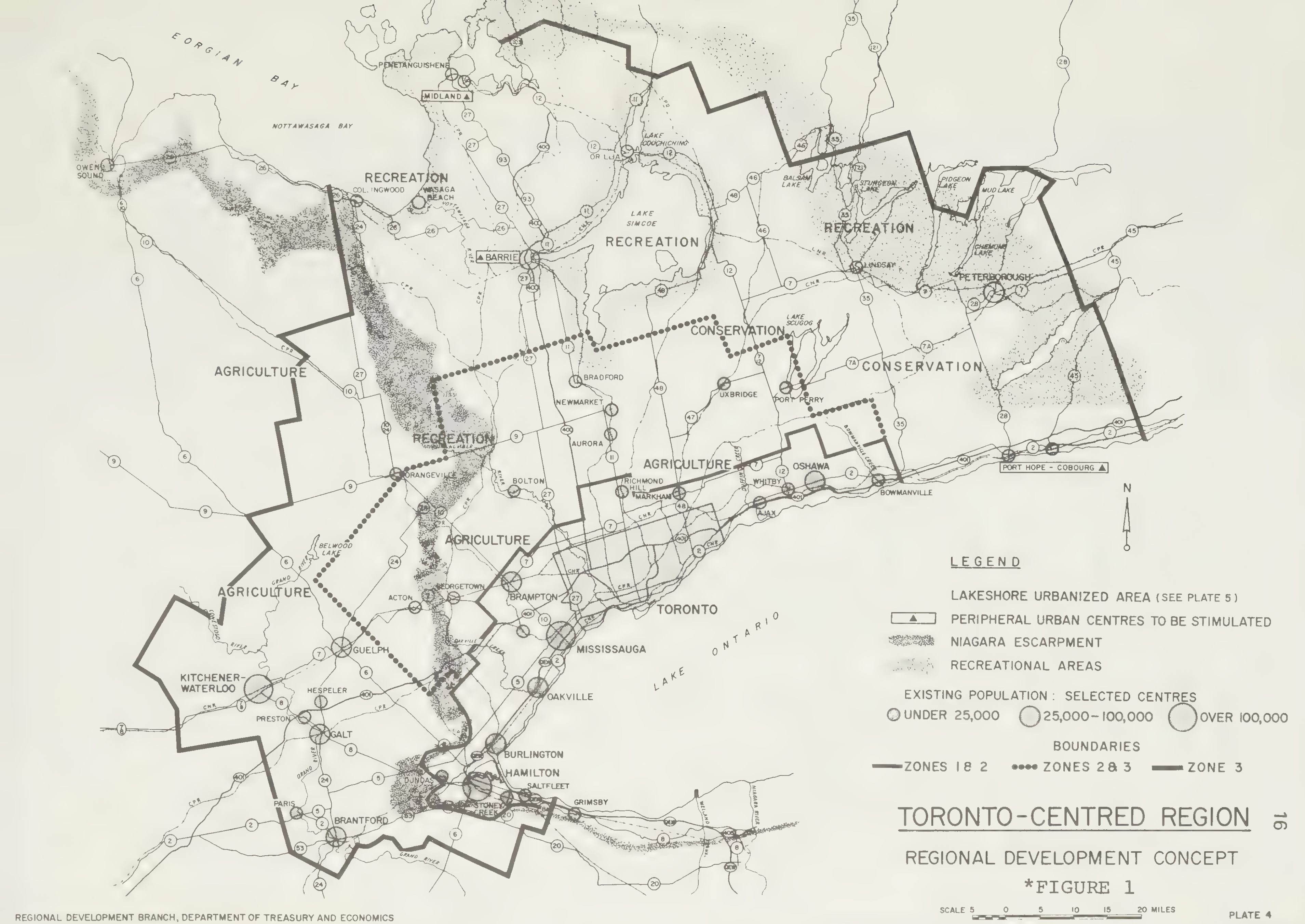
For non-urban areas, thorough coverage is especially important to the Area Development Agency for purposes of detailed identification and conditions of economic stress.

Recommended grid cells were related to population density. Then we gave a list which indicated our desire for:-

(a) Total coverage, and

(b) An increasingly fine mesh as the population density became more pronounced, both present and expected.

I have not changed my views too much since that time. I do believe that this type of information is needed, and very definitely on a small area basis. Indeed when I came to the province, we were working with small units and one of our immediate problems was to establish some kind of common denominator assessment of change, social and economic. We therefore, set up a rather ambitious program of examining changes according to the smallest geographic unit for which data were obtainable. We have some 63 indicators of social and economic change, which will be published before the end



16

*FIGURE 1

*Page 16
Design for Development: The Toronto-Centred Region
Government of Ontario, 1970

of the year as a Regional Development atlas. We were able, through using these, to set down formally what many of us knew intuitively, that there were parts of the province which were performing well, parts of it performing intermediately well, and other parts not so well. We can, therefore, tailor our policies to this type of objective information. In practice, we found ourselves hampered time after time by the fact that data were available all too often at only the county mesh, so our maps have an element of crudeness in this regard, which I think is unfortunate but which Geocoding could overcome.

Finally, we went into the actual regional planning. We found ourselves once again with this problem. I have asked that there be circulated among you, (and I take it that each of you now has) a copy of the Toronto-Centred Region for reference. I would refer you to page 16 (see figure 1) of the report, to the map of the region. Now I am sure you are familiar with this concept, which was released last May 5th, and was the result of several prior years of effort. It is a concept, at the provincial level, for the growth of Toronto, and the region within a general arc of 90 miles from Toronto. In the course of our deliberations it became necessary to separate a highly urbanized corridor which we call Zone 1, an intermediate low-density area which we call Zone 2, and which is just north of Zone 1, and an outer area which we call Zone 3 which extends from Kitchener-Waterloo to Peterborough and includes Port Hope and Cobourg.

Now we were faced immediately with problems of how to draw these zonal boundaries in such a way that we could get data on the one hand and incorporate the major policy ideas on the other. I should say, at the outset, that basically we are encouraging structuring, or restructuring of growth within Zone 1; particularly conservation, open-space maintenance and agriculture in Zone 2; and stimulation of growth, particularly in the northern and eastern sections of Zone 3, while structuring growth in the Western section. The purpose is to make the fullest possible use of the Toronto-Centred Region.

We finally came down to the township when drawing these boundaries. You can see that there are certain unusual lines that we have some difficulty defending publicly. The Prime Minister of Ontario has said that this is to be considered by departments as a guideline and, therefore, certain reactions are filtering down, and in our follow-up meetings we are being asked such questions as 'Why did you draw that line here, or there and so forth'. We offer answers that are as plausible as possible, but I think you can see that, in certain instances, it would have been much easier to have drawn those lines if we had had the proper Geocoding procedure. Now, within the next thirty months it will be necessary to refine this concept on the basis of various suggestions that have been brought forward to us and we are under a fairly strong pressure from certain areas to make certain amendments to these boundaries. On what data, you see, do we make these? We are getting all

types of data from vested interests, but that is not quite what we feel we should use in the final analysis. So we do have a problem here. We think it can be resolved in the interim, but I am just using this as an example, where if we had had a really thorough Geocoding procedure, we would have been a little farther along. I would use this as a key example of how we can benefit by much of what this conference is going to present today.

In summary, therefore, I would say we urgently need Geocoding. We need it in standardized grid units, such as I am sure will be proposed today. In my view, there should not necessarily be a universal grid of the same mesh, but the grid should increase in fineness as the complexity, not necessarily the population density, but the overall complexity of the situation increases or is expected to increase over years to come. It should be possible for the finer meshes to aggregate into the coarser ones. I should say, as the need arises, we need to consider not only the static information which is gathered in the normal grids, and I am sure Mr. McDaniel will get to this one more in detail, there is the all important matter of flow phenomena, which are very critical. So these are some of the key questions which must be evaluated, certainly at the senior technical level. In Geocoding, we are looking to this technique to provide the answers.

Thank you very much.

Mr. J.I. Weldon
Coordinator, General Survey Systems
Dominion Bureau of Statistics

"BASIS CONCEPT AND GEOCODING BY DBS"

Thank you Mr. Chairman. I should like to get down and discuss this morning the so-called nuts and bolts of Geocoding. I will be describing the development of Geocoding in simple terms which should be understandable by those non-technical people who are primarily concerned about what it can do for them and how it can be used.

The problem of Geocoding came about, after the 1961 Census, when we published the census tables. The information was disseminated and the census data were assessed by people. It was realized that a census is usually coded to so-called pre-determined standard areas; in census context this usually means a census tract, an enumeration area, municipality or province. Most of our tabulations, or a large number of them, are aggregated at a census tract level.

Stated more precisely, the problem is this: data are collected according to fixed areal units, whether they be political as with census track, counties or municipalities, or administrative as in traffic or school zones. These data, practically speaking, can only be retrieved by the particular political or administrative unit in which they were collected. Thus, if data are collected by census tract, it is not possible to obtain data for an area smaller than

a census tract, for example half a census tract or 2 2/3 census tracts. It is possible to aggregate whole census tracts. However, we should realize that the census tract is not the only way of aggregating data over a particular geographical unit and that there are many different interested groups, who do various administrative development work, such as planners, financial people, economists, city engineers and so on, who all have their own pre-defined administrative work areas. For example, the planner may be interested in some sort of a planning zone; the economist may be interested in different types of economic regions; city engineers may be concerned with traffic flows or traffic patterns. These examples serve to illustrate that different groups of people have their particular reference areas, and it is almost impossible that these various reference areas should be conforming close enough to the census tract, so that census tabulations will be meaningful to these particular reference areas.

This is why the concept of Geocoding was developed. We said, wouldn't it be nice if someone could delineate any area in the urban region and say, give me some data for this particular urban area. And then, after assessing the tabulation, say, "It wasn't quite the area I wanted because the tabulations don't bear it out, so I'll change it slightly and now can I have some data on this other area?" So the keyword was flexibility - to be able to get information for any particular area.

Now the word 'Geocoding' means the assigning of geographic co-ordinates to a location, a location which is represented by an address - as in 33 Douglas Drive. Why geographic co-ordinates? The purpose of geographic co-ordinates is to provide the basis for the retrieval of information. The query area (area for which information is requested) can be geometrically described as a polygon (see figure 1) and defined by the corner points, the so-called vertices. A polygon can be viewed as a series of straight lines. There is a simple geometric, mathematical algorithm from which we can select all the points which lie inside of this query area. Knowing the vertices of the polygon and knowing that all points are referenced to a geographic co-ordinate framework, it is possible to automatically select by computer all those points within this delineated area.

The co-ordinate reference framework which has been adopted after a considerable amount of soul-searching and studying, is the 6 degree UTM system, the Universal Transverse Mercator System. Geocoding really means that to each data observation there is assigned a co-ordinate reference point, in the UTM 6 degree reference system. If we visualize a tax assessment file and each record in this file contains an address plus a string of data, then by Geocoding the address, we are, in effect, adding the XY co-ordinates to the location in this spatial reference framework. Once someone delineates a query area, the computer selects all the records in the file which


FIGURE I
LONDON MAP
DEFAULT OPTIONS AT 2000 SCALE
CREATED OCT 04 69
AT 2204 HOURS
OPTIONS SPECIFIED
Scale=2000 ft/in=1:24000
Features = Name
Centroids not plotted
No node symbols
No address ranges
St. signs not plotted
Control points not plotted
St. dir'ns not plotted
Overlap
St. begins not plotted
Character size = 0.050 in.
PLOTTAPE IS VOL SER L00501
Map prepared by DBS
TYPICAL
QUERY
AREA

*FIGURE 2

MUNICIPALITY		CODE	TYPE	NAME		FEATURE	DIRECTION	DESCRIPTION		ERROR
3500	1948	18400	ST	TIMOTHY				ADDRESSABLE FEATURE		
SEQ	NODE	X-REF	TYPE	X--CO-ORDINATES--Y		LEFT	RIGHT	X--CO-ORDINATES--Y		
060 E	011415	PINE	ST	308,680 308,607	4,879,129 4,879,130	623	618	308,578	4,879,075	
055	011414	HOLLA		308,630	4,879,115	599				
050	011413	CHARL	ST	308,549	4,879,089		568			
045	011412	PROSP	ST	308,463 308,337	4,879,064 4,879,048	543				
040	011411	C N R		308,332	4,879,024					
035	011410	HOLLA		308,250	4,878,997	497				
030	011409	CEDAR	ST	308,225	4,878,991					
025	011408	MAIN	ST	308,182 308,071	4,878,977 4,878,964	481 431	480 436	308,085	4,878,923	
020	011407	CHURC	ST	307,975 307,889	4,878,911 4,878,906	413 383	422 384	307,902	4,878,865	
015	011406	JOHN	ST	307,817 307,722	4,878,861 4,878,853	377 333	378 334	307,735	4,878,812	
010 B	011405	LORNE	AV	307,641	4,878,805					

*Area Master File for town of Newmarket
Produced by Department of Highways and
Ontario Statistical Centre

have spatial reference co-ordinates within the delineated area.

The important consideration in this deliberation, is our ability to convert addresses into co-ordinates. I will emphasize urban areas, because most of my remarks this morning are about urban areas. There are two types of Geocoding, urban Geocoding and rural Geocoding. But I am going to talk in the context of urban Geocoding. Typically, urban areas consist of some sort of rectangular street pattern, and each street segment has two block faces, and each block face has an address range, for example, from 568 to 618. To geocode, what we can do very simply (figure 2) is look at a certain address, let's say Timothy Street #580, and look up our reference files of block faces and see to which block face on Timothy Street this can be allocated. We find that it is the block face whose co-ordinate value is 308,578 - 4,879,075.

This block-face centroid is a XY co-ordinate which represents the centre point or approximately the centre point of a block face. Once we have found which block face the particular address fits into and, since we have already determined by previous work the co-ordinates of the block face centroid, we can then add to the address the appropriate X,Y, co-ordinate. Now, I can add this then to my data file. It must be realized that all the other households or properties, which are in any particular block face, will get the same XY

co-ordinates. Thus, when a query area is delineated, the Geocoding retrieval system will search out all the data records which have XY co-ordinates within the delineated query area.

The advantage of this system is that we can delineate an area of any shape, within limitations. The limitation, specifically in the urban area, is that the query area should be delineated along street lines not across street lines. The retrieval algorithm is not that sophisticated, to deal with these marginal cases.

Another nicety about Geocoding is that we can produce plotted maps of the urban areas, and we have put on the wall a couple of these maps (figure 1). These maps are really a by-product, recording all the block face reference co-ordinates in the city.

I should, in all fairness, point out that you can do a lot of good work even without going into Geocoding, that is determining block face co-ordinates and assigning these block face co-ordinates to every address. I'm particularly referring to the good work that the City of Vancouver has done, quite a few years ago. They capitalized on the nice regular layout of the street pattern, the hundred block system, and they are able to delineate the area by streets and which are defined by street intersections for example, 8th Avenue and 54th Street. They have a programme capability to retrieve everything inside these city street boundaries.

However, Geocoding has certain advantages, first

of all because of mapping capabilities, and secondly, it is less error prone than other methods of retrieval. Even the City of Vancouver decided some time ago to convert their system over to the Geocoding way. The Americans are doing Geocoding also, but in a different way; they will be able to provide census tabulations from census files, on the basis of listing all the blocks for which census tabulations are required. Now obviously, if you have a large city such as Toronto, in which you can have several thousand blocks and you have to delineate a large enough area, there is a lot of coding required to list several thousand city blocks which constitute the retrieval area.

The Geocoding development, briefly, with no technical terms, can be described in three stages, bearing in mind that what we wish to have is some sort of an address conversion file, as I have indicated, which will list all the block faces by street name, by address range of the block face, the low and the high address range of the block face and, of course, the XY co-ordinates (figure 2). This is what we want to get out. The work, however, is a one-shot effort and, once you have built this conversion file, the only work remaining is the up-date, to effect any further changes in the street pattern. The work consists of building an area master file (the address conversion file is a subset of the area master file) which describes every street in terms of street names, the block face in terms of street

intersection co-ordinates, address ranges, and then automatically generating the intersecting street names of every intersection and the XY block face centroid. This work is fairly time consuming. My rough estimate would be that it will probably require two man months clerical work to produce the input preparation, say per 100,000 population, plus or minus 100 percent. We are just completing the Geocoding for the City of Toronto and it went quite well, something in the neighbourhood of 8 months for about 4 people; that would be about 32 man-months, and Toronto has 2 million people so it's within the plus or minus 100 percent range.

After the area master file is produced, which essentially describes every block face in the city, together with block face co-ordinate references, there can be derived several special purpose application files. The one which I have already mentioned is the address conversion file, which describes every block face by street name and address ranges, and XY co-ordinates. Another file which could be created is the input for plotting the street map. A third application might be just to produce a street index. This is essentially the stage 1 work to implement Geocoding.

We now start the application for Stage 2. There is available a file, which might be the census file, the 1971 census file or your Municipal Assessment Roll file. To perform the Geocoding retrievals, that is to retrieve information by user specified areas, one has to assign the XY

co-ordinates to each address in the assessment file. The concept of the address conversion is quite simple, and yet I would like to go through it again just to be quite sure that everyone appreciates it. The address conversion file contains street names, address ranges and corresponding XY co-ordinates, which is sorted alphabetically by street names, and within street names, by house numbers. After the sort, it will be a simple merge routine of all these street names, then all addresses for a particular street within a block face will be given the same XY co-ordinates. It is a very simple mechanised automated operation. It works quite well but there are problems. One problem may be that we haven't produced the address conversion file accurately - there may be clerical inputs or new developments since it was produced, requiring updating corrections. Another problem might be that the addresses are not properly spelled or wrongly defined. There is a problem with the address itself, even the Municipal Assessment Roll file for the Province of Ontario has 37 characters set aside for the address. Aside from the fact that the first five digits are house numbers, we don't really know for sure which characters contain the street name, apartment name, city name and so on. In other words, we can say that this address file is sort of a semi-free format, street names have different lengths, and therefore we can never be sure just how many characters will be contained in the street names. We have, therefore, developed a so-called

accuracy decoding programme which accepts the designated area, the 37 characters, as the area which contains the address in unknown format, and then this programme proceeds to determine which portion is the house number (this will be easy as they are always written in the first five digits) which portion is the street name, apartment name, municipality name and whatever else might be in the address. After this has been determined, the address is then put into a standard form, that is a certain position always for the street name, (no matter how long the address) and so on. We have now a formated address file, which can then be passed against the address conversion file. By doing this, we can then transfer the XY co-ordinates from the appropriate block face listed in the address conversion file to the address of the data record.

When the data file is geocoded, that is the XY co-ordinates are assigned to every address, the only task remaining is to store the file in suitable format and retrieve. Our objective was, as far as retrieval from the 1970 census file was concerned, to be able to retrieve any combination of census data characteristics by any kind of user specified area quickly, cheaply, and in good turn around time. Anyone who has had dealings with DBS in the past would have been frustrated that the turn around was less than desirable. It typically ran into several months, or sometimes even longer. To aid in turn around time, we have developed, therefore, in

accordance with the specification, a special purpose retrieval programme, which has several features. First of all, due to the randomness of the application, (any city or any part of a city in Canada) the census file is stored in a random access file, that is on magnetic discs. Random accessing capability and file design has enabled us to compress the file and we figure that the 1971 census file, which was originally on a hundred reels of tapes, will be reduced to around 25 discs. Because of random accessing capability, turn around time will be faster. We are aiming at and so far have maintained, overnight turn around.

There was another retrieval problem which had to be overcome. In the past, not only at DBS but in many data processing centres as well, to retrieve information from a file, a retrieval programme had to be written. The programme typically is quite simple, it takes a few days to write, then, after it is written, it has to go through testing, debugging, acceptance testing, user verification and so on. All this takes time and these were mainly the reasons why anyone having asked in the past for any special purpose tabulation from DBS had to wait several months. What was required was a generalised programme, a generalised programme that the user himself could learn in a few hours time, even if he had no previous computer experience. We have developed this generalised programme which has a limited English

dictionary. You can express your request, your query in an English-like constructed manner close to the problem orientation on hand and write it up, and this request then is directly the input for the retrieval programme. It is key punched, submitted and hopefully, your results should come back overnight.

Every request for census information must refer to a query area. Every request potentially could refer to a different query area. So, coupled with the retrieval capability which I have described, someone must delineate the query area. DBS, therefore, is proposing to provide maps, street maps, probably around 2,000 feet to the inch (similar to one illustrated in figure 1). The user then would look at the map, presumably with a red pencil to delineate along the street line the retrieval area and he would send this map in with his request. Suppose he wants a tabulation by age, sex, marital status, income. The delineated retrieval area will be then the boundaries, the corner points would be digitized by the operations group and this would be an extra input to the request.

In the first step, the retrieval will determine which block faces are in the query area, which are data points and which households are within those block faces. It will retrieve those data points, and then perform the tabulation for those retrieved data points.

Another feature that we have developed is inter-

facing the retrievals with the SYMAP programme. SYMAP is simply a programme which produces computer printed maps rather than plotted maps. Because we have already a Geocoded file with all the XY co-ordinates and retrieval capability for any area, the interface with SYMAP programmes can be accomplished with just a few days of preparative work where you just write in effect a normal retrieval request and a few additional things.

This essentially covers my discussion of the conceptual aspect of urban Geocoding. As far as the non-urban Geocoding is concerned, in the rural areas, there are no street patterns, street addresses or house numbers, sometimes you don't even have streets. In rural areas the urban Geocoding concept is not going to work. It was mentioned earlier this morning that 80 percent of the population in the next decade will be urban residents, but we still need something Geocoding-wise in the non-urban areas and, as far as DBS is concerned, for the census file we will represent each enumeration area by an XY co-ordinate. We will have essentially the same capability in non-urban areas as in urban areas. If someone can delineate the query area on a rural map, we will digitize the boundaries of the query area and retrieve all the data within the enumeration areas which are within the prescribed query area.

In the other part of my talk today I would like to briefly describe the census plans for the 1971 population and housing census, conducted by DBS. This decennial census,

which is going to be held next Summer, will enumerate the entire population; the 100 percent enumeration will be restricted to information about sex, marital status, age and a few minor questions in relation to the household. Also, every third household will be given a much expanded census questionnaire, roughly 120 questions will be asked, various social, economic, migration, labour force, fertility type questions. Because of the large universe, a 1 in 3 sample, should produce quite accurate census tabulations. As far as the Geocoding aims are concerned, we have been geocoding for the last year and a half, the twelve largest cities in the country. Going from coast to coast, these are Halifax, Quebec City, Montreal, Ottawa, Toronto, Hamilton, London, St. Catharines, Winnipeg, Calgary, Edmonton and Vancouver. We are not necessarily going to geocode the entire census metropolitan areas, or even the entire municipality in the legal sense. It is a large undertaking and there was limited time and limited resources. We will always capture the most densely populated portion of the city and that part will be geocoded, but it depends on the resources and availability of good maps and address ranges. In some places, we do cover pretty well the entire city, in other places we don't cover quite the entire city.

Once this work is in progress and, by the end of this year, we should have finished the area master file preparation and, therefore, the address conversion file

preparation, for all these cities, and when the census data come in, after having it edited and processed, we are going to assign the geographic XY co-ordinates, the census address, and then the file will be stored on a random access device for flexible retrieval.

We hope that in the operation there will be some users service group in DBS which will act as the liaison between the users, public and computer processing and statisticians in the Dominion Bureau of Statistics. Hopefully, users who will have frequent requirements from census tabulations, might accelerate the process work by learning the retrieval language, but it is not necessary. They can describe what tabulation, income, etcetera, they want from the census file. They would have to obtain and mark up a so-called query area map and they would delineate the query area. This would all be sent to the users service group in DBS which would then determine the query areas, the co-ordinates of the boundaries and then submit the retrieval tabulations.

I would like to say a few words about query areas itself. There are, that I can think of, several types of query areas. I started out my talk with the most common one which is the predetermined, predefined or standard query area. This would be census tracts, city areas, municipalities, for which we even have the appropriate codes in the file.

The other type of query area which is probably more

important to you people here, is what I call administrative areas, for instance, a set of planning zones which constitute all the planning districts in the city, or the neighbourhood districts. The engineering department might be interested in having permanently declared and defined all the traffic zones; they might want to get various different types of tabulations by traffic zones, and instead of defining every time when a tabulation is submitted, this could be defined once and for all and then you can just refer to the traffic zone by name, Traffic Zone 1, 5, 16, 19, etcetera.

Finally, we come to the third type of query area, called the'ad hoc' specified query area. Typically, perhaps a graduate student wants to study various aspects of economic development and he wants to do various steps in the dark, and he specifies a query area with the understanding this is a one-time retrieval, we will have to digitize it, go through the process of defining all the data points within this area and he may like it or he may not like it, he may modify this request by respecifying the query area.

I am quite hopeful that, in as much as census tabulations are concerned, you will experience a major improvement in turn around. We are still aiming at a few days turn around, and I hope it will be within a week or a couple of weeks elapse between sending in your request and returning your tabulation.

This completes my discussion of Geocoding.

Thank you very much.

PROFESSOR R. McDANIEL
Department of Geography
University of Western Ontario
London, Ontario

"GEOCODING IN RESEARCH"

Mr. Schnick, colleagues, ladies and gentlemen, I will begin my brief talk this morning by providing some perspective on Geocoding.

Geocoding is, in a sense, a natural evolution of an interest in location which began very early in the age of astronomers, geographers and navigators. Anyone who has ever been interested in location, for navigational purposes or for the purpose of locating oneself on a map, has in a sense been a forerunner of Geocoding. Geocoding is an evolutionary phase of mapping, brought about through information technology.

More immediate predecessors of Geocoding might be found in aircraft plotting procedures, introduced during World War II. Currently, when we look around for models that may be at least analogues with the types of things that may emerge from a full-fledged Geocoding system, we might consider weather maps, and such exotic hardware as the SAGE or NORAD defense systems. Perhaps many of you, if not all of you, have seen pictures of these electronic display systems showing maps with the locations of aircraft and various defense installations. The trend, which is beginning with an interest in Geocoding, would appear to take us in a direction which might be construed

as a civilian extension of this military system. There are many examples of new technology evolving in a military environment and subsequently being applied in a civilian environment. The computer itself was, in large measure at any rate, an outcome of World War II when the concern was for rapid computation of shell trajectories, and least cost aircraft trips. The SAGE system is a very sophisticated system, but it is still primitive when contrasted with the likely civilian applications which may emerge.[1]

Essentially, we are interested in analyzing what may be termed time-spatial phenomena, that is the analysis of patterns over the surface of the earth, and through time. Weather maps, as mentioned, are an example. We can produce a time series set of digitized weather maps, run the data through a cathode ray tube, and actually see the simulated weather move over the surface of the earth. With any Geocoded data on a time series basis, we could process it similarly. Some geographers have experimented with such computer mapping procedures and observed a city's population changing and spreading amoeba-like over the landscape. Research applications of such capabilities might be rather limited, but certainly from a pedagogical standpoint, it is extremely helpful for students to actually perceive the varying rates of differential changes in the spatial pattern. From a research standpoint, it might be argued that considerable insight into the processes involved might be had from being able to actually observe the animated pattern. As you can see from the maps, particularly the gridded maps on the wall, with the possible

exception of the sign map, the Geocoded road system appears much like a network or a graph. I find it rather interesting that concurrently with this development in Geocoding there is growing interest in the general area of network analysis. Such a Geocoded road system would enable the determination of least cost flows through the network.

As Mr. Schnick mentioned, our department at Western has recently completed the construction of a Geocoded data bank for at least parts of the Lake Erie region.[2] There were several disciplines involved in this within our department, that is five sub-disciplines of Geography involving an economic geographer and an historical geographer. Each of us contributed Geocoded data to this particular bank. Now the size of the area which was Geocoded varied with the context. In the case of the physical data, for example, it was decided to adopt the 500 metre grid square, and all information in it, based upon an analysis of topographic sheets, was generalized. Such things as soil type, soil texture, drainage and other information for a given square were attached to that particular code number (the military map reference of the centre of the square).

The agricultural data were coded on an individual parcel basis. That is, a farm might be composed of several parcels, and in a rather detailed analysis of Yarmouth township, it was noted that the majority of farms comprised several spatially separated parcels, and each parcel was assigned a geocode which again was the centre point of the parcel. Each owner was given an identification number so that it was possible to assign each

Geocoded parcel to a specific owner. These were subsequently mapped and it is rather interesting to observe the variation in distances among these spatially separated parcels.

The economic data, for which I was primarily responsible, differed substantially in the nature of the geocode used. I might mention at this time that there are two very broad systems of Geocoding, which might be labelled 'Location Coding', and 'Naming Coding' on the other. Location Coding requires the assignment of the actual geographic co-ordinate, latitude and longitude. XY co-ordinates are actually assigned to a point or an area, or to a line of a block face. The 'Naming' procedure, which at the present time is much more common, involves attaching a number, very arbitrarily, to a location, usually a political sub-division. Thus a census tract might be given a number 15; if that census tract is within a city which might be numbered 900, then that census tract in that city would be 915. If the city is in a province, then attach another number in front of the 915, perhaps 7, to identify, say, Ontario. Because of the global distribution of the locations with which I was concerned, I restricted myself to the latter type of coding, that is the 'Name' variety.

The data which confronted me, unlike those of my colleagues who were using data strictly limited to the Lake Erie region, pertained to material coming in to the industries and various other economic activities in the region, and materials going out. I was interested in the origins and destinations of these

various inputs and outputs. When interviewing an official of a firm, we sought to know the various materials consumed by that firm, and specifically from what point these originated, and similarly, for their products, we wanted to know their destination. Origins and destinations were name-coded by city, province, state or nation as given by the firm.

The disadvantage clearly of this so-called 'Name' system is that one cannot compute distances among points. However, for some frequently used data, perhaps a naming system might complement the Geocoding system, although as has been pointed out this morning, the Geocoding system is clearly much more flexible, and could enable one to detail any particular area. Thus, most of my comments regarding possible research applications of Geocoded data will pertain to Geocoding of the Location type, which permits one to compute distances.

But, in spite of the limitations of the hierarchical Name system which was employed in the economic sector, we were able to begin at least tentative exploration of the effect of spatial aggregation upon certain measures. The basic accounting model used, and some of you may be familiar with it, was input-output analysis. An input-output matrix of some 42 industries was computed for each of the four counties in our region, and for the largest urban centre in each county. We were able to sequentially aggregate the data of these spatial units, and observe how sensitive were the technical co-efficients (cents worth of input per dollar of output) to the variation in the degree of spatial aggregation. We also examined the proportion

of total inputs to manufacturing from a given area and noted how this changed as we aggregated the areas. Thus we computed for the manufacturing sector of Woodstock, the proportion of total inputs from Woodstock, then from Woodstock plus Oxford county, and so on for ever-increasing areas to all Ontario, then all Canada. We did this for both inputs and outputs for the total Woodstock economy, for the retail sector, as well as for the manufacturing sector, primarily to test the retrieval capability of our system.

From a purely research standpoint, we were interested in exploring the effect of spatial aggregation on various measures. All of us are very much aware that certain measures change as we aggregate sectors of the economy and similarly, as we aggregate spatially, certain measures change. However, without a Geocoding system we are severely restricted in the flexibility in this kind of experimentation, which is needed to determine the degree of sensitivity of our economic measures to spatial aggregation.

The discussion of input-output analysis introduces the category of analysis to which Dr. Thoman alluded earlier, that is the analysis of flows. The emphasis so far this morning has been on the stock or static data pertaining to block faces. Another major category would be flow data. Thus we would like to have, not only the Geocoded static data on population for certain cities in the 1971 census, but also Geocoded origins and destinations of migration flows, commodity flows, labor and shopping flows, recreation flows, information flows and money

(capital) flows.

Now the ideal must abstract obviously from the problem of privacy, and this is a problem with which we are clearly confronted in simple sectoral aggregation, the aggregation of firms in order not to reveal information on a single firm. There are comparable problems, perhaps more difficult to resolve, I dare say, in spatial aggregation. While I talk about the desirability, from the ideal standpoint, of knowing the flows from individual firms in one specific location to individual firms in another location, this probably will not be forthcoming, but we might be able to aggregate by certain areas, and by certain kinds of flows.

A colleague of mine, who is studying urban phenomena, tested his data by attempting to define the urban edge of London, on the basis of density of certain functions. The weakness of this kind of study involving density is that it depends on the unit of area which you are going to use. Once you define an area, however, say a square mile or some other standard unit of area, then with a Geocoded system you can quickly determine the number of units in each of these areas. If you then define the edge of an urban area in terms of the density of activities, then you have a more or less objective tool. I say more or less, of course, because of this question of defining the area. However, this method provides a way of comparing, fairly objectively, the growth, shape and size of many cities, and perhaps contrasting and noting discrepancies between this definition of city area, and that based upon actual municipal boundaries.

The Geocoded system of data recording will clearly enable us to very quickly observe various spatial patterns. I find it difficult to think of the Geocoded system in terms of a once and for all reporting. Once we have the kind of technology, which has been discussed here today, generally accepted, and once electronic accounting systems are fairly general throughout business, and in accounting firms, we can expect data to be forthcoming much more quickly. Thus the system mentioned earlier, of a time-spatial series of maps for analyzing the dynamics in a spatial system, would certainly be feasible.

There are a number of practical applications of such a system in the context, I think, of transportation. Given a Geocoded transportation map, we could determine mathematically the distances in a highway network, or an airline net, or a railway net, and could complement this with a computerized traffic rate file, instead of leafing through these voluminous manuals on traffic rates. The Geocoded data file could provide us with the basic data for the determination of the route, and the transport rate file would provide the cost, and merging these two would enable very quick determination of least cost transportation routes for practical or purely academic experimentation purposes.

Returning briefly to the density of activity concept, one could think in terms of interrelating such kinds of data with other data. If one is interested in testing some hypotheses of sociologists that urban density affects human variables such

as mental health, we might be able to explore the implications, and such implications as are supported could then be fed back into planning decisions.

Generally then, there are a number of possible research uses of Geocoded data, and I might conclude by recapitulating;

1) observing the development of an urban shadow phenomenon actually over time, in different areas;
2) studying the time distance effects of changing transportation lengths and changing time factors, such as might arise through the introduction of new technology;
3) observing variations in various flows over space, or assessing the impact of future flows upon the system;
4) observing both static and flow data changes as they occur in something approaching real time, if we define real time as time consistent with controlling or influencing the process in question;
5) sorting out areas or places with a certain set of characteristics for study, so if you are interested in studying certain kinds of industries, certain kinds of people, you have a convenient sorting procedure for finding out where they are, and you can very quickly zero in on where the areas for research exist.

Summarizing then, I see a Geocoded system, such as discussed here today, as a precursor of a much widely expanded, much more dynamic system which in many respects will replace the maps with which we usually work. As a final thought, I leave

you this possibility of reducing our concern with drawing boundaries on a map. We often draw boundaries on a map for administrative convenience because we do not have detailed information, and the information we do know is based on political divisions; thus we tend to draw lines around these political divisions. There are other reasons we want to draw lines on a map, of course, such as for allocating responsibility for new kinds of points which may enter the system. However, in fact, as we are very much aware, there is much overlapping of mapped regional responsibilities among departments. This leads to strait-jacket-like attempts to develop common mapped regions. But most of these departmental problems, if we reflect upon it, pertain to specific individuals at specific points, or to specific activities at specific points, and if we could deal with these specific points, these specific individuals, rather than simply arbitrarily drawing lines around them, we may have a much more flexible system. A Geocoded system of data collection may provide the statistical basis for an effective co-ordination of government departments, business firms and other social institutions as a smoothly functioning total system.

References

1. See an elaboration of this point in H. Sackman, Computers, System Science and Evolving Society, Wiley, 1967.

2. R. McDaniel, A.K. Philbrick, V.W. Sim, H.W. Taylor, M.J. Troughton and W.R. Wightman, The Erie Research Project: Construction of a Multi-Purpose Data Bank, Department of Geography, University of Western Ontario, 1970. (A Report to the Regional Development Branch, Ontario Department of Treasury and Economics.)

Mr. D.C. Weeks
Project Leader
Electronic Computing Branch
Department of Highways

"IMPLEMENTATION OF THE DBS GEOCODING SYSTEM IN ONTARIO"

(Editor Comment - This presentation was accompanied by illustrative slides and a demonstration of information retrieval by computer).

If we go up in the tower at Niagara, we can look down and get an overall picture of the city below. We can see how one thing is related to another. The astronauts went up even higher and saw how whole countries were related to each other. If we could get this kind of information into the computer, we would have a very powerful tool.

Most of us here are government people. In order for us to serve the public, we must know where the need is. When ships are out of sight of land, they use the geographic co-ordinates, latitude and longitude. This system has been known for a long time. But we have so many other ways of telling where we are that the co-ordinate or XY systems have fallen into disrepute. However, it is still the only universal way of locating things.

Some people may be wary of getting tangled up in geo-co-ordinates. They think the user will have to deal with large X and Y numbers. In reality, the user will never see the co-ordinates.

In some cases, we should not try to adapt old referencing systems to the computer. XY co-ordinates are very compatible with computer operation.

Old referencing codes can be added to the new system.

Another way of looking at it is that XY co-ordinates can be added to an existing file. This is not as hard as it sounds because the new system can automatically add XY, if the address is known.

Many people are building geocoding systems. This will allow the computer to retrieve stored information in small areas. However, these systems may not work well together. The exchange of information between agencies may not be possible.

Some people may choose to leave the development work to others. This offers the advantage of being able to profit from others' experience. Of course, an agency could take someone else's system and adapt it to their own needs. However, such a modification should be agreed to by the original authors of the system. In the case of such a modification, the basic systems would remain the same and would probably be compatible with each other.

The Dominion Bureau of Statistics has developed a system for pinning XY on various items. At the request of the Ontario Statistical Centre, the Department of Highways of Ontario is obtaining the DBS Geocoding System. The Electronic Computing Branch of the Department of Highways is now implementing the DBS System for use in Ontario.

The system is divided into three parts. The first part is the master file phase. This phase sets up the street network in the form of XY's. The larger Canadian

cities have already been recorded in this manner. This includes Toronto, Ottawa and Hamilton.

In the master file phase, the XY co-ordinates are established for each street intersection. One side of the street between two intersections is called a block face. The centroid of each block face is arbitrarily found by offsetting the mid-point between intersections by 22 metres.

The second part of the system, the geocoding phase, will analyze the street address accompanying incoming data and will automatically store the data with the proper block face centroid. By the way, in case you're wondering what a meter is, it's a device used by the utility companies to help bill you.

The part of the system which will be used most is the retrieval phase. This phase can tabulate information from any existing file. We have distributed some sample retrieval results. The retrieval language does not require an input sheet, it is entirely free format.

The keywords must start in column one of the input cards. There are 6 key words: filename, areaname, heading, selection criteria, characteristics and tabulate. These names can be shortened to their first letters.

"Characteristic" is where you name the details you want. Such as age, sex, cars owned, floor area of house, etcetera.

"Tabulate" allows you to define the type of table you want as output.

"Selection Criteria" or "S" allows you to restrict the records you want to examine to certain broad categories, such as farmers' daughters.

"Filename" or "F" is the name of the information file you want to question. This file must be reduced to a bunch of funny names before you can use it.

That is, you must tell the computer what names you are using for the different items in the file. This is done only once, when the file is first established. The people who want to question the file must have a list of the funny names you have used.

To find out how many farmers' daughters there are in the file, you could write this:

```
C:  ES7    'FD';
T:  COUNT;
```

ES7 is the funny name for electoral status.
(instead of the word characteristic, we can write "C"). This may look strange at first. But the pattern is always similar and you can learn it quickly.

We can cross-tabulate to find the number of times certain combinations of details occur. For example, how many house owners have 1 car, how many have 2. The request could be written like this:

```
C:  ES1   'O';
    CARS  '01', '02';
T:  COUNT;
```

The codes are in the same list as the funny names. The table produced would look like this:

	CARS	'01'	'02'
ES1 '0'		---	---

The funnynames and codes appear as row and column headings. Many dimensional tables are possible, for example, by further categorizing the house owners and tenants according to age and sex.

Of the six keywords, the one which characterizes this as a geocoding system is "Areaname". This allows us to search only within a selected area. We can define the area by drawing the boundary on a map or by naming the streets enclosing the area. The system will locate any user-defined area, not just standardized areas like census tracts.

Many people collect information. However, sometimes it is hard for other people to get at the information. So they go out and collect it again. People get angry after the hundredth canvasser comes to call to get the same data.

If there were a common form of reference, the information could be collected once and used by everybody. Service organizations could provide data more easily if there were a common denominator. Almost all data we work with can be related to physical location. Therefore, it is logical that XY could be this needed link.

To sum up: should we use the system? Answer: we should use some sort of geocoding to know where things are. Should we use this particular system? Well, why not try it. It's very easy to use. Also, it is backed up by the Federal Government.

It provides some ready made free data, in the form of geocoded street networks. Since we are basically noble, how could we repay the debt. By helping to record other cities in XY co-ordinates. They will be useful regardless what system we finally end up using.

We might also help by making some of our files available to others. But, you say, we feel responsible to our citizens not to threaten their privacy by releasing information to others. The problem of personal privacy is very real, and must be reconciled with the good we can do by knowing the facts. One way we can protect individual privacy is by using larger retrieval areas, and averaging the results. Another way is by asking the people what facts they would not mind releasing for the community good.

Geocoding is a big concept which can be a big help. It would be well to keep in touch with what is being done in the field of geocoding. This would help us keep our new systems compatible with the idea of geocoding. If possible, we should start using geocoding soon to gain first hand experience. That way, the computer will also be able to get the overall view, and we will profit by its knowledge.

GEOCODING SYSTEMS IN THE UNITED STATES - 1970

A General Overview With
References to Canadian Experience

This paper has been prepared by

Edgar M. Horwood
Professor of Civil Engineering
and Urban Planning
Director of the Urban Data Center
and the Urban Systems Laboratory

and

Charles E. Barb, Jr.
Assistant Director, Urban Data Center

University of Washington
Seattle

Note: The production of this working paper was supported by the National Science Foundation under Research Grant GS 2832 being conducted currently at the University of Washington. The opinions expressed are those of the authors to whom comments and criticism should be directed and are invited.

Professor Edgar M. Horwood
Professor of Civil Engineering and Urban Planning
University of Washington
Seattle

"GEOCODING SYSTEMS IN THE UNITED STATES - 1970"

For the purposes of this presentation geocoding is generally defined as a systematic process for deriving geographic codes for data entities which are identified by street address. Whether the input comes from census data collection efforts or other sources is immaterial; geographic codes may relate to standard type areal units defined by the census authority, other standard areas of local utility or _ad hoc_ areal units defined in fact after the data collection efforts.

Even within the definition stated above and the qualifying assumptions, there are many different efforts underway in the United States and Canada, to the extent that communication is becoming difficult and a new vocabulary is in the making. An ancillary purpose of this discussion is the presentation of a simplified taxonomy of these systems and explication of terms now commonly used.

In spite of rather substantial efforts to standardize geocoding on a national basis by both the U.S. Bureau of the Census and the Dominion Bureau of Statistics, local systems are developing on their own in many urban areas. Most of these are being designed to meet local needs or both local needs and census related summaries. As might be expected, the significance

of this tool to local users is of such importance that many applications cannot await the development of a national standard system. As an example, the mandate of the U.S. Supreme Court on school integration is requiring hundreds of school districts to make detailed analyses of pupil location and apply greater thought and analysis to the network routing of their buses. Under this kind of pressure and with the requirements to minimize the costs of acquiring large fleets of buses, Orange County, Florida, commenced its own geocoding system over a year ago.

As in the case of computers themselves or programming software the issue with geocoding should no longer be dependent upon national, provincial or state standardized systems. It may be functional, indeed, to have a duality of systems in metropolitan areas to serve both local and national needs and users. Parts of systems will be interchangeable and interagency agreements may expedite this exchange. In this regard the classification of geocoding systems and development of vocabulary will be of great importance.

A CLASSIFICATION OF GEOCODING SYSTEMS

The classification system presented here is best discussed with reference to Figure 1. A geocoding system basically consists of a directory, now commonly called in the U.S. a Geographic Base File (GBF), and a directory access system, or search procedure. The product of geocoding is the assignment of geographic codes or co-ordinates to street addresses. After the assignment of these spatial designations additional soft-

Figure 1

A CLASSIFICATION OF
GEOCODING SYSTEMS - 1970

	GEOGRAPHIC BASE FILES						ASSOCIATED FILE ACCESS SYSTEMS	
	FILE CLASS BY BASIC ENTITY	FILE PURPOSE	TYPICAL X-Y DESIGNATION WHEN PRESENT	TYPICAL NETWORK DESCRIPTION	RELATIVE FILE SIZE FOR URBAN AREAS	GENERAL EXAMPLES	NAME	ACCESS METHODOLOGY
ADDRESS CONVERSION PRIMARY TO SYSTEM	Street Length Intercept	Directory	Area Centroid	Absent	n/4 (n equals one city block)	U.S. Census Street Index - - Regional Area Trans. Study Directories		Traditionally Manual
	Block Face	Directory	Block Face and Adjoining Block Centroids - - Block Face End Coordinates	Absent	4n	U.S. Census Address Coding Guide (ACG) (no coordinates) - - NYC System (with coordinates)	AD-MATCH	Serial Match
	Street Segment	Directory	Block Centroid - - Segment End Coordinates	Implicit	2n	Seattle GEOBASYS Directory	GEO-BASYS	Random Access
						U.S. Census DIME		
ADDRESS CONVERSION ADJUNCT TO SYSTEM	Parcel	Data Base	Parcel Centroid - - (Perimeter coordinates as content data)	Absent	20n	NCC System Ottawa Region		

ware and hardware are necessary to make these assignments meaningful in applications, such as by creating summary tables, contour maps, dot maps, etcetera. These latter systems are essentially post processors of the geocoded output and will vary widely with the multitude of applications.

Geocoding systems may be most appropriately classified according to their directory type, of which the four principal ones are shown in Figure 1. The following discussion will be structured according to these four directory types.

Street Length Intercept. The most common of these are the street indexes relating address ranges to census tracts. The streets of the directory are listed in alpha-numeric order with the address range intercepts allocated to each census code. Both the U.S. census Bureau[2] and the Canadian Dominion Bureau of Statistics[3] have produced or have encouraged the production of this type of directory. To date, such directories have been used almost entirely in the manual assignment of address coded data to census tracts or traffic analysis zones in the case of the urban area transportation studies. As an example, every day in Seattle an official of the local health department hand codes birth and death certificate data to census tracts.

Directories at this gross level of abstraction have rarely assigned geographic co-ordinates to the area codes. If assigned, however, either through hand measurement or automatic digitizing processes, hachure, dot, or contour maps can be constructed. As an example, in the typical SYMAP process the

approximate centroids of census tracts are recorded and the summarized values of data for the tract are assigned to that location for subsequent derivation of contour shaded areas by the computerized process.[4]

Street length intercept directories typically have far fewer records than the other types shown. The typical census tract area file in the U.S. will have about 25 street length intercept records for an area including about 100 city blocks, or approximately one record for every four blocks. Although attention is now shifting to more refined file structures, these will contiue to be highly serviceable for certain applications and otherwise for urbanizing land areas lying outside of the confines of the more extensive systems. With automation they may, in fact, constitute a second level order geocoding system to be used for applications which do not require the refinement of block face level specificity.

Block Face Directory. These directories are in fact a refined modification of the type just discussed. Rather than constituting an index of street lengths within a large statistical area, these directories link the street name and address delimitors for each block face to geographical code at the block level as well as upwardly through heirarchial coding. In fully automated systems the block face centroids as well as related block centroids are commonly assigned x-y coordinates for subsequent production of dot map or contour map displays.

Considerable interest in this type of directory was engendered by the Address Coding Guide (ACG) program of the U.S. Bureau of the Census undertaken with the co-operation of other federal and local agencies in 1968. The ACG's served as an input to our first national direct mail enumeration effort.[5] To support development of the ACG's, the Geography Branch of the Bureau of the Census developed an entirely new Metropolitan Map Series based on the U.S. Geological Survey Quadrangle Maps. The working scale of this series is mostly 800 feet to the inch, with some portions of the inner city areas at 400 scale.

The ACG program, however, does not go as far as digitizing any of the geography.[6] This is a point of confusion because the much publicized New Haven Census Use Study which did digitize geography in what has now come to be known as the Dual Independent Map Encoding (DIME) system.[7] At the time of this writing, the majority of the larger metropolitan areas in the U.S. as well as some of the smaller metropolitan areas have had their ACG's recoded to include digitized census geography. This recoding was conducted through the "ACG Improvement Program" and results in creation of DIME files. This recoding has similarly been effected through co-operative effort of Census and other federal and local agencies. Unfortunately, this current recoding process being conducted in the U.S. does not include thorough editing or updating of the 1968 base map, the street names, address ranges, or census codes.[8] As a consequence, at the completion of the program, most local areas will be far from having quality (i.e.

operational) geocoding systems. At most, the ACG and ACG Improvement Programs will provide a substantial input for potential geocoding systems; at least they have kindled local interest.

The other side of the coin, however, is the raising of false hopes among many that operational geocoding of local utility will come about either without much more local effort or at an early date, or both. A number of informed and experienced observers as well as ourselves question the wisdom of the census in upgrading the ACG, which we believe is essentially a poor start given the ACG's current unedited condition. Our experience in the ACG and ACG Improvement Program in the Seattle area implies strongly that the upgrading of the Census products may be more time consuming and costly than starting de nouveau and using modified on-line editing procedures upon a Cathode Ray Tube (CRT) for both the verification of the street geometry and coded information.

There is no doubt that the ACG's can constitute the directory part of geocoding systems of ongoing utility to a local region if ongoing editing and updating procedures can be introduced. The directory is only one part of the geocoding system, nevertheless, (and unfortunately) many local areas in the U.S. believe they will have operational geocoding systems when they receive their ACG directory from the Census. Actually, they are many thousands of dollars, many organizational headaches and many years away. Even the much publicized New Haven

Census Use test system, developed for the most urbanized portion of that urban area, has slipped into oblivion. No ongoing operational capability has been established locally or remotely with the Bureau to provide local service.

Street Segment Directory. This type of directory is philosophically very similar to the block face file. The major difference is the reduction in the number of file records by approximately half through the use of the street segment as the entity (i.e. information for the adjacent block faces lying on both sides of the street segment is included in the same record).

Another difference not yet broadly appreciated about the U.S. Census DIME file is the identification of the street network as a separate network to that formed by the boundaries of census blocks. A surprising number of census blocks contain boundaries that are not part of the street system or their boundaries do not account for discontinuities in the street network caused by abrupt changes in grade level or the superimposition of grade separated freeway networks. The Census Bureau's DIME system can be modified to distinguish street continuity from census block boundaries, but geocoding systems are not that far along by and large to have this requirement emerge. The importance of developing a street network of ground truth accuracy lies in the use of network algorithms for many studies that are dependent on street time/distance solutions. One of these, the Moore tree building algorithm,[9] also permits the summarization of data for the data entities connected by the

tree network such as that for population, housing and school attendance.

Street segment directories are highly utilitarian for local studies even without census code conversion capability. Such a directory, constructed at the Urban Data Center,[10] has been used in the Seattle area for several years. 1970 Census codes are now being appended to the directory to give it the same coding capabilities as the ACG and DIME directories. Our research suggests street segment files will become increasingly important for traffic engineering applications, bus routing studies, street department accounting, etcetera.

Parcel File Directory. Much attention is now focusing on parcel file as assessment data become machine readable. These are somewhat functionally different from the types previously cited, in that address conversion to geographic codes is an adjunct use of the file and in fact may be done more economically by the use of other types of GBF's. A good example of this type of file is found in the assessment records of Hull, Quebec, under development by the Canadian National Capital Commission (NCC).[11] The use of such files for spatial analysis is being inhibited in the U.S. by assessment file development procedures. Current U.S. assessment file development procedures are designed to deliver tax statements to the appropriate owner or lien holder and do not commonly integrate the street address of the property as a data item.

While the parcel file is highly important, its size

may limit its effectiveness as a general purpose geocoding data delivery system. At typical urban densities there are in the order of 20 land parcel records for each block, or about ten times the number of records that exist in the street segment file. We believe the parcel file has an important use in geocoding by facilitating the conversion of property related data to spatial determinants, but that its direct geocoding function is limited. On the other hand, if directory updating and maintenance becomes formally lodged in the assessing function, the problem of directory maintenance may become effectively resolved. We see these files as separate but equal ones to those previously discussed and worthy of continued exploration.

Directory Access or Search Systems. The second basic element of a geocoding system is the directory access or search system. As Figure 1 documents, current systems include the traditional manual look up approach, the serial tape-to-match approach (ADMATCH),[12] and the random access approach (GEOBASYS) we are pursuing at the Urban Data Center. We feel the random access approach, while possibly no faster than the serial approach, will point toward future real-time applications of geocoding individual addresses.

Comments upon the U.S. Bureau of the Census ADMATCH program at this time are still only tentative. The system reportedly has been installed upon only sixteen machines and our examination of it has been only limited.

THE EMERGING POLITICS OF GEOCODING SYSTEMS MAINTENANCE

The geocoding systems to be delivered through the U.S. Census effort in 1971 and 1972 will still be based on maps of 1968 vintage and will be at some yet undetermined level of accuracy. Studies we are now making in the Seattle area indicate that levels of Census GBF accuracy will vary from 70 to 90 per cent in the central city and may go as low as 60 per cent in the suburbs. These kind of results will probably disappoint many prospective users.

The plain facts of the matter are that directory maintenance and updating are very expensive processes, even excluding the costs of maintaining search and display software. Under these circumstances, there will surface many consultants who will be able to assist an agency in address matching but only to a level of accuracy inherent to the Census product. With the Census Bureau's sale of ACG and DIME directories to all comers at marginal reproduction costs, there will be a highly competitive environment for applications consultants and little impetus for public agencies to close in on the maintenance updating problem. The Census Bureau has in fact pre-empted local experimentation, inhibited reasonable budgeting and destroyed the incentive for local co-operation toward producing quality and useful geocoding systems. Fundamentally, we question whether highly technical and location-specific geocoding systems can adequately function except under local control and upon a proprietary basis even in the public sector.

CONCLUSIONS

There is all too little understanding of either the operational environment, hardware and software requirements, and user assistance institutions needed for geocoding systems. At the present state of development and with the emergent outlook for Census sponsored systems, considerable user anticipation exists. Unfortunately, users will only become fully aware of the Census product limitations by 1972 or later, and reliable geocoding systems will not become operational in most urban areas until the late Seventies. The delay of these systems and experimentation with their use will keep viable urban information systems in the horizon for some time to come. Local dependence on Census products may preclude realistic budgeting at the local level to meet the problems discussed, and, more important, delay or preclude the organizational development of sound regionwide and unified directory building and maintenance institutions.

FOOTNOTES

1. The Orange County Project was reported by Dr. Gordon Foster, Director, Florida School Desegregation Consulting Center, University of Miami, Miami, Florida, during a conference entitled "Computer Applications to Desegregation", 5-7 November 1969, at Florida State University, Tallahassee.

2. The U.S. Bureau of the Census directory is commonly called a Census Tract Street Index. See U.S. Bureau of the Census, Census Tract Manual, Fifth Edition, Washington, D.C., U.S. Government Printing Office, January 1966, pp. 51-55.

3. The Canadian Dominion Bureau of Statistics, equivalent to the U.S. Bureau of the Census for many functions has for some time prepared and distributed equivalents to the U.S. Census,Census Tract Street Index for major cities in Canada.

4. The U.S. Bureau of the Census "Census Use Study" has researched and documented several computer mapping systems including SYMAP. See: U.S. Bureau of the Census, Census Use Study: Computer Mapping, Report No. 2, Washington, D.C., U.S. Government Printing Office, 1969.

5. Address Coding Guides have been developed for all 233 Standard Metropolitan Statistical Areas in the United States (by definition, counties in which cities of larger than 50,000 population exist). For a list of these areas and the Census geocoding systems available for them see: U.S. Bureau of the Census, Census Use Study: The DIME Geocoding System, Report No. 4, Washington, D.C., U.S. Government Printing Office, 1970.

6. New York City has created essentially an Address Coding Guide with geographic co-ordinates independent of the U.S. Census. Creation of the file is most recently described in Herzer, Ivo, "Case Study: Creation of a Geographic Base File", published in Papers on the Application of Computers to the Problems of Urban Society, 5th Annual Urban Symposium, August 31, 1970, New York City, Association for Computing Machinery, 1970.

7. Op Cit, Census Use Study: The DIME Geocoding System.

8. The editing effort was expressly limited in a preamble to the ACG Improvement Program Supervisor's Manual. The stated objective of the program has been limited to making the ACG and Metropolitan Map base map merely to agree.

9. Moore, E.F., "The Shortest Path Through a Maze", published in The International Symposium of Switching Proceedings, Harvard University, April 25, 1957. Cambridge, Harvard University, 1957.

10. Several documents have reported the Urban Data Center's research of the street segment type directory. The research has been supported by the U.S. National Science Foundation since 1962. Pertinent references include:

Dial, Robert B., Street Address Conversion System, Research Report No. 1, Seattle: Urban Data Center, University of Washington, 1964.

Calkins, Hugh W., Operations Manual for Street Address Conversion System, Research Report No. 2, Seattle: Urban Data Center, University of Washington, 1965.

Crawford, Roger James, Jr., Utility of An Automated Geocoding System for Urban Land Use Analysis, Research Report No. 3, Seattle: Urban Data Center, University of Washington, 1967.

Barb, Charles E., Jr., "Street Address Conversion System" a summary description published in Papers from the Sixth Annual Conference of The Urban and Regional Information Systems Association, September 5-7, 1968, Clayton, Mo., John E. Rickert, Editor. Kent, Ohio: Kent State University, 1969.

The prototype Seattle Street Address Conversion System (SACS) is currently being supplanted by Seattle-King County GEOBASYS managed by a consortium of local public agencies. GEOBASYS will have a county wide directory and a third generation production directory access system. GEOBASYS is scheduled to commence operation in January 1971.

11. For information concerning the Canadian National Capitol Commission System contact Mr. David C. Symons, National Capitol Commission, Ottawa, Canada.

12. ADMATCH is currently being distributed by the Central Users' Service, U.S. Bureau of the Census, Washington, D.C. Price $60. A systems user manual is available for purchase separately for $0.75.

Mr. D.C. Symons
Chief of Computer Services
National Capital Commission
Ottawa

'RURAL AND SINGLE PROPERTY GEOCODING'

Thank you, Mr. Chairman, Ladies and Gentlemen.

We have been experimenting with Geocoding in the National Capital Region and I would like to talk, this afternoon, about what we are doing. We have completed Geocoding a city of 60,000 in the region and have started parcel geocoding the Ottawa-Carleton part of our region. We are building an information system for urban planning, and our multidisciplinary group of experts were brought together specifically to build this information system for planning and analysis in the National Capital Region.

Under the information system, we have five basic sub-systems, of which parcel Geocoding is one. The first sub-system is data collection, or the data base. This is based in part on municipal assessment and census records. All of these data are updated annually and the information from some communities is updated quarterly. Data standardization across the region varies considerably between Ontario and Quebec, thus, we have a bilingual area in which to test our Geocoding system.

The second sub-system is the retrieval system and

we are presently experimenting with generalized file management, with an English-like command language. We have tried GIS and are looking at other data management systems.

The third sub-system is graphic display, similar to what Dr. Horwood discussed, except that greater emphasis will be placed on production of line drawn maps.

The best part, and perhaps the most interesting, is the model system. This uses data retrieved through parcel Geocoding. Of course, this system takes into account transportation studies, economic models, land use models, and other forecasting techniques.

Parcel Geocoding requires three basic things, in our view. We need an automated assessment and municipal file, base maps showing ownership parcels tied in with the modified Transverse Mercator system, and capable software to manage the data.

We have recorded all the data from the assessment file. It was decided to record the centroids of each ownership parcel in the city of Hull, approximately 60,000 population, as a test case. We elected to record the block corners because this allows us to define the streets as a parcel. We are really interested in covering all the land area in the municipality, not just the private ownerships, and very frequently, large ownership parcels are not covered in the assessment file. City parks, for example, may not appear.

We are recording 100 percent of the land and water area in the community. By digitizing the block corners, we end up with a street segment between intersections. Because we have taken off the other blocks, we end up with a description of the intersection. Thus, we record everything, including the co-ordinate boundaries of all the properties. Simple!

But, in practice, it is a little more difficult because adequate base maps are necessary. We must have tax mapping and assessment maps tied in to the 3° Transverse Mercator grid.

Our retrieval system is based on the fact that we wished to be able to draw an arbitrary configuration on the map and retrieve the data points that fall within it. We have seven polygon point retrieval algorithms and various combinations: for example, the intersection of two polygons, the intersection of a circle and a polygon, the intersection of a vector and another polygon, additional subroutines, and so on. The advantage of this type of system is that we can build a location directory of a vacant property, which obviously does not have a civic address, we can create one. It is possible to produce a location directory of commercial establishments, banks or any specific uses we may wish to identify, provided the data are contained in the data base.

As a matter of convenience, we carry a minimum of three grid co-ordinate values on each map sheet so we can

properly scan each map section. Block numbers are also recorded to facilitate retrieval. A few users will not specify arbitrary polygons, but will request blocks 1, 2, 3, 4, 5, etcetera, and list these. We estimate that to capture all the data for the property boundaries has added about 20 percent to our cost for digitizing. It required approximately two-man-months for the co-ordinate digitizer to take off the co-ordinate values for the city of Hull, roughly 15,000 assessment parcels provided accurate base maps with roll numbers are available. Considerable time was saved because the editing was done visually. The data was output on an incline plotter, edited on the plot, and corrections were fed back into the system.

The evaluation of rural areas uses the same software user for the urban areas, except that the block-and-parcel may be reduced to the 200-acre-lot-and-concession system which is prevalent in that part of the region.

In the National Capital Region the lots and concessions were laid out in the original plan about 150 years ago. The smallest lot is about 200 acres, so we used this and we digitized the centre of each 200 acre lot. If we did not have individual ownerships, we simply assigned subsequent subdivisions to this larger entity. In addition, we record the co-ordinates of the corners of the concessions and rural roads so we can identify ownership and data along the roads and provide a location directory. The same software is used in both

urban and rural systems. We tested this system but have decided to use individual ownership parcels in rural areas.

We have excellent compatibility with the Dominion Bureau of Statistics census because the Geocoded data carried in the data base may be summarized by any arbitrary polygon and, therefore, the system can retrieve data for any traffic zone, census tract, or enumeration area we may wish to specify. I would like to emphasize that this flexibility is built into the system because we carry the data in its disaggregated form.

We have talked about a number of advantages of the system. We are not tied to civic addresses, so the system may be extended to include non-urban data. The transitional area, between the urbanized portion of the region and the agricultural area, is where the development will occur. We expect to be able to measure the changes in urban development for small areas on a yearly basis.

The beauty of this system is that the computation is simple. The data retrieval is based on a point-polygon algorithm. This simple algorithm requires not more than 5K core and is quite flexible. It may easily be extended to include additional methods such as the union of two polygons.

We are experimenting with the location of municipal services and public utilities. It is easy to record the co-ordinate locations of invert elevations of sewer lines, and

the locations of power poles and the like. We expect, by the Fall of 1971, we will be able to retrieve information about municipal services and public utilities from the data base.

With the co-ordinate values contained in the assessment file, that we have entered on the assessment file, it is quite easy to put out a plot, of a typical assessment map.

It should be noted that for each of the line intersections shown on this map we carry two seven digit co-ordinate numbers. This is a typical map sheet which we digitized, the lower left, the lower right and the upper right show the three degree MTM control points. The numbers shown for each parcel with the asterisk indicate the file number. These are transverse mercator grid co-ordinates. The streets and intersections are shown on the map. Also, we have a scale option in the program which allows us to change the scale.

The two large parcels that were not contiguous on the map sheet are indicated as a broken block. Rural Geocoding would look like the pattern shown on this drawing. The scale option, one inch equals 75 feet, may be varied as needed. On the maps from which we have been digitizing, one inch equals 50 feet.

I would like to deal with a number of problems that are, perhaps, outside the scope of a discussion on Geocoding,

first is data standardization across the region. We find there are great differences in the way data are recorded. The second problem is that adequate mapping is frequently not available in many parts of our region, although fortunately we have a survey group attached to our organization and we simply went out and asked them to map in those areas that weren't mapped. I wouldn't recommend anyone starting on this type of a system unless they are prepared to go and do integrated mapping for the urban area. As a practical matter, adequate mapping is being done frequently by many different agencies and my recommendation is that co-ordination should be considered.

We are extending the system to include other data such as the location of employment, to be keyed to the location of residence. We expect that we will be able to originate yearly origin and destination of travel matrices in the region for this particular community on a test basis. The problem is not the limitation of the Geocoding system but rather the limitation in the standardization of the data base. We know this works, but we need the data to operate.

Thank you.

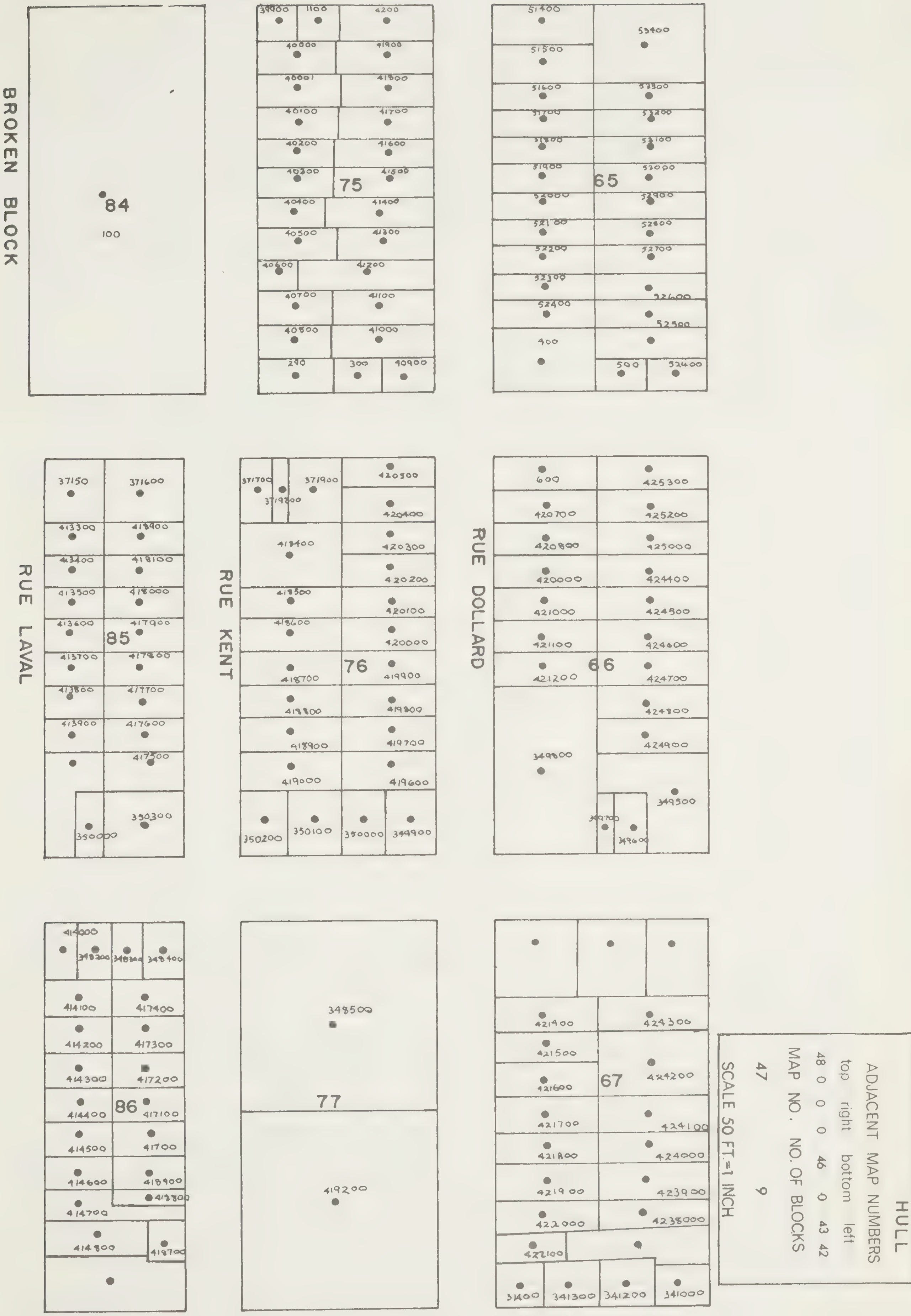

BROKEN BLOCK
RUE LAVAL
RUE KENT
RUE DOLLARD
84
75
65
85
76
66
86
77
67
HULL
ADJACENT MAP NUMBERS
top right bottom left
48 0 0 0 46 0 43 42
MAP NO. NO. OF BLOCKS
47 9
SCALE 50 FT.=1 INCH

FIGURE 6


100
90
101
81
92
HULL
ADJACENT MAP NUMBERS
top right bottom left
41 0 49 48 43 0 35 0
MAP NO NO OF BLOCKS
42 5
SCALE 75 FT=1 INCH

FIGURE 9

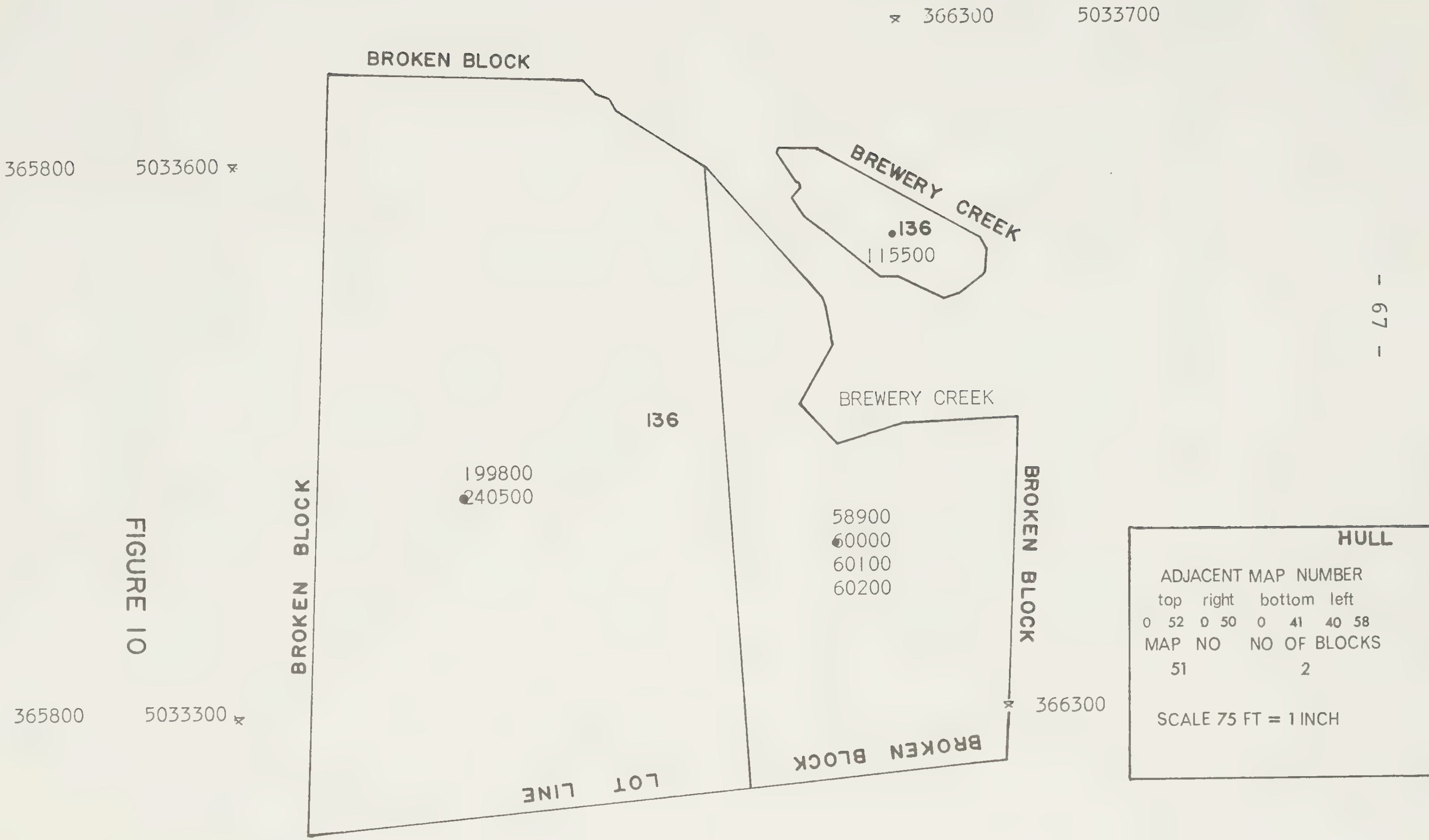

366300
5033700
365800
5033600
365800
5033300
366300
BROKEN BLOCK
BREWERY CREEK
.136
115500
BREWERY CREEK
136
199800
240500
58900
60000
60100
60200
BROKEN BLOCK
BROKEN BLOCK
LOT LINE
BROKEN BLOCK
HULL
ADJACENT MAP NUMBER
top right bottom left
0 52 0 50 0 41 40 58
MAP NO NO OF BLOCKS
51 2
SCALE 75 FT = 1 INCH

FIGURE 10

Dr. K. Cheng
Director
Ontario Statistical Centre
Department of Treasury and Economics

'GEOCODING ACTIVITIES IN THE ONTARIO STATISTICAL CENTRE'

Thank you Mr. Chairman, Ladies and Gentlemen. Today we have heard some excellent talks on the various aspects of Geocoding. I would like to continue in that same vein and relate to you a few of my thoughts on this subject. However, I will first begin my remarks by discussing the geocoding activities that have been undertaken in the Ontario Statistical Centre.

Geocoding has been of some interest to us for several years as a statistical tool to aid in the collection, storage and retrieval of information. Our first major plunge into the geocoding waters was in 1969 when, with the co-operation of DBS and the City of Hamilton, we initiated a project for the geocoding of the metropolitan area of the city of Hamilton. This project has now been completed. The data for the computer retrieval demonstration, which you saw this morning, came from the test area of this project.

In 1970 we negotiated with DBS for the transfer of their geocoding programs to the Ontario Government. With our assistance, the computer programs are now being installed at the Department of Highways (who, by the way,

have shown keen interest in this field). Coincident with this, we are engaged in the geocoding of the town of Newmarket, using Department of Highways facilities. At present, our geocoding enthusiasm is limited only by our financial constraints.

As has been mentioned by Professor Horwood, the geocoding system most commonly employed in the United States is called the DIME system. The American DIME file and its Canadian counterpart, the Area Master File, are quite alike in purpose and design in that the basic building unit of each system is similar, that each system will allow for the production of computer printed maps showing the street network, and that retrieval of information is by any area configuration. The basic unit of the DIME file is the street segment while, with DBS system, the basic unit is the block-face. This means that there are fewer records in the DIME file as in the DBS system. The reason being the street segment consists of both sides of the street while the block-face represents one side only. It is my thinking that the DIME system, by employing the 'block chaining edit' approach, may have a better editing procedure than the DBS geocoding system.

The demonstration of data retrieval, which you have witnessed today, had shown that data can be retrieved by many different characteristics as age, marital status, sex, religion, etcetera. Data retrieval can also be by any specified area within the geocoded area. It is my

thinking that a great advantage of geocoding will be in the area of file linkage. In the future, data files can be linked with one another, such as a health file linked with the population Census file or municipal assessment with Census or other administrative files. It will prove useful and feasible to information users to have information from these various data files cross-tabulated. This linking of files could be accomplished with the aid of geocoding by employing files which already have been geocoded (i.e. addresses replaced by a x-y co-ordinate). The linkage could be accomplished through the use of a common identifier which, in this case, is the block-face centroid. You will remember that the block-face centroid is the x-y value which is attached to each address in the particular block-face. Since the centroid value is the linking agent between files, it is necessary that the x-y centroid value be the same as between different geocoded files[1]. In the case of single property geocoding, file linkage could be accomplished through the property centroid. I believe that this type of

---

[1] If, for a particular geocoded area, there are files which, for one reason or another, have different x-y values for the same block-faces, it will still be possible to link these files by re-geocoding one of the files. This process only involves computer time, the cost depending upon the length of the file.

standardization should be encouraged by the Ontario Government and Dominion Bureau of Statistics to ensure the successful development of an effective information retrieval system.

In any discussion on geocoding, it is important to give consideration to the type and quality of data which will be geocoded. Technically speaking, any information which is location specific and which has an address can be geocoded. However, there are practical limitations to the type of information files which can be geocoded. If a data file is not in suitable machine readable form (that is, on computer cards or magnetic tape), it cannot be geocoded. The cost of record design, coding and keypunching of some files, may be so high as to render its geocoding impractical.

Besides being in machine readable form, the file must be reasonably 'clean'. A 'clean' file is one which contains what it is supposed to contain and nothing else. It has been our experience with our two geocoding test projects in Hamilton and Newmarket that the data files have been considerably lacking in cleanliness. With the Municipal Assessment file for the town of Newmarket, we found that almost 20% of the records were missing, either partially or completely, the property address.

Any file which is being considered for geocoding, should contain information which has been standardized. It is impossible to have any meaningful retrieval of information

with unstandardized data. For example, in our Hamilton test area, in searching the occupation field for the number of physicians, we didn't know whether to search for the name doctor, medical doctor, M.D., physician or Dr.

I have one final comment on data files. Some of you may find this point obvious but sometimes the obvious has a way of being overlooked. It is extremely important that there be provided adequate documentation about the data file. There must be a good description of the record layout, field format, together with a dictionary of terms in the file. It has been our experience that, in this regard, there is considerable room for improvement.

When we are deciding which files to geocode, considerable expense and headaches can be avoided by giving strong consideration to those files which have been properly developed and maintained. It is in this way, with clean data files, that the geocoding capability can be fully exploited.

Most of the discussion today has centred around the type of geocoding which is applicable to urban areas in which the only types of files, which can be geocoded, are those which contain a civic address to represent the location of the data. While this system will operate quite effectively in large urban areas as Toronto or Hamilton, where the civic addressing system is comprehensive, I feel that, as one applies geocoding to smaller and smaller urban areas,

certain problems will increasingly arise. The source of the problem will probably be an incomplete addressing network. For example, in our pilot project of the town of Newmarket, we have found that, of a town of 10,000 population with approximately 5,000 assessable units, there were at least 10% of properties without complete addresses and, as such, were not readily codable by our machine. This 10% comprised vacant lots, businesses in the main thoroughfare, and houses. The usual situation was, instead of 103 Main Street, we would find in the data record N/S[2] Main Street. I am pleased to report that the majority of the imperfections have been overcome. With the aid of up-to-date civic address maps, provided by the town of Newmarket, we were able to identify vacant lots and give them a pseudo address, also we identified most addresses which contained a N/S or E/S numeric code and we were able to give them an appropriate number address within the suitable address range.

For rural areas, that is areas without any type of civic addressing, it has been argued that we need a different system of geocoding - a rural geocoding system. There are merits to this suggestion but I think that we could make do with the present system, for the meantime anyway. I think that it is possible to have a rural geocoding system which

---

[2] N/S meaning North Side

will use the urban geocoding concept as a framework. For rural areas, instead of the civic addressing system, we could have a pseudo type addressing system which would take into account lot and concession type numbering, instead of having the block-face as with urban areas. The basic unit would be the lot. The lot centroid could replace the block-face centroid.

There is much scope for development in this area of rural geocoding, particularly to refinement of a system to assign co-ordinates to each lot.

In any geocoding system, whether urban or rural, single property or block-face, there is a constant need for updating. This updating of the Area Master File should take into account changes in urban environment as changes in boundaries (as with Newmarket), changes in existing street patterns, address ranges, street names, new housing developments, etcetera.

From the information point of view, the geocoding of data is not only useful for provincial and federal governments but also for local governments and agencies. Since these local governments may wish their particular municipal area to be geocoded, it will be desirable, both from the economical as well as the operational point of view, that they be requested to participate in the creation and updating of the Area Master File.

As to the matter of cost of developing the Area

Master File for any urban area, the answer is not clear cut. It is not possible, for instance, to say that your Area Master File will cost so many dollars per block-face. What the cost will depend upon is availability and quality of basic geocoding inputs. For instance, if a particular municipality has good street maps, at proper scale, a comprehensive and up-to-date civic addressing system, accurate control points, accurate street index, then the cost of creating the Area Master File is what I feel quite reasonable.

Assuming the availability of inputs, such as just mentioned, the cost of Area Master File creation for an urban area of population 50,000 people would be in the neighbourhood of a couple of thousand dollars. This estimate does not include the cost of geocoding a data file. That cost will depend again upon many variables, particularly length of file and degree of cleanliness.

In closing my remarks, it will do well to point out that Geocoding is not some magical statistical instrument that will give us reams of data at the push of a button, but rather it is a technique for the storage and retrieval of information from existing files. How useful Geocoding will become in the future will depend to what extent we, meaning the various levels of governments, agencies and departments, are willing to co-operate with one another in an effort to develop an integrated geocoding system in the Province of Ontario.

Thank you.

SUMMARY OF PANEL DISCUSSION

Dr. Cheng opened the panel discussion with a few introductory remarks.

A question was asked regarding the applicability of geocoding (i.e. planners' information needs are different from those of engineers).

Mr. Horwood replied that his own philosophy is that regarding applications, there is no need to consider separate systems for different information needs, whether private or public. The street segment method does allow for the artificial creation of street segments (i.e. where the particular block is densely populated or unusually long).

Regarding the accuracy of the file, it is Mr. Horwood's opinion that a system that can have its accuracy increased by means of an easy editing process is one that will pay off with more applications. In other words, if we were to take that file and show it at a scale of about fifty feet to the inch and compare it with an engineer's map, we could correct our centroids through the online editing process for that section to match any accuracy we wanted. Mr. Horwood believed that there was no justification for two separate systems as long as the editing process is capable of upgrading the system to any degree of refinement. In this manner one would ultimately meet everybody's needs.

A question was asked regarding the XY coordinate system and whether in the future there will be a need of a Z

co-ordinate as well.

Professor Simmons responded that in his work he intends to add a Z co-ordinate to the file at a later date. The use of the Z co-ordinate for land use in built-up areas is quite important, where one might want to show the land uses on the ground floor, and above ground floor, or on different floors. Also condominiums, air-rights and things of this nature, further point to a need for an additional co-ordinate.

Another questionner commenting on the increasing geographical orientation of our data basis asked whether we are getting anywhere by having better geographical data, if we don't have better geographical statistics. In the past, in using unit areas for the storage of data, people have noted that, if census tracts are used, there is a certain correlation of certain variables but if enumeration areas are used, there is a different correlation.

Professor McDaniel responded that the problem is intertwined with our concern for privacy, while Mr. Horwood stated that we should differentiate between sensitive data and location data.

Mr. O.M. Schnick
Executive Director
Economic & Statistical Services Division
Department of Treasury and Economics.

'CLOSING REMARKS'

Ladies and gentlemen, it would be inappropriate to conclude our session without making certain remarks about the day. First of all, we have had an attendance of some 100 persons, a broad representative group of senior personnel who are interested in the subject of geocoding, and we appreciate very much your presence here today.

During the morning session, it was apparent that the various speakers were well integrated in relation to the subject at hand. Mr. Weldon explained to us what geocoding involves and how his division is implementing the system for the forthcoming population Census. Professor McDaniel has given us an insight into research applications for geocoding, while Dr. Thoman gave us a comprehensive talk on the problems associated with the gathering of data on a small area basis. With regard to the computer retrieval demonstration, I think I am correct in saying that it was most interesting. Mr. Weeks has given us a good explanation of how a general retrieval programme functions.

After lunch we heard from Dr. Horwood on the subject of geocoding in the United States and Mr. Symons has given us a clear insight into rural and single property

geocoding. Dr. Cheng has informed us of the Ontario Statistical Centre's activities in geocoding. The extensive use of geocoding in the United States and the application of this technique to the forthcoming 1971 Canadian Census would seem to indicate that government personnel should give some serious thought to the merits of this system. This brings us to the purpose of the seminar which is to initiate an assessment of the need for geocoding in Ontario.

Many of us present are employed by the Ontario Government and may be potential users of any geocoding system. Consequently, in approximately two weeks, a letter requesting your comments will be sent to the various government departments.

In closing, and on behalf of all present, I would like to thank the Cafeteria Staff for an enjoyable luncheon. Our thanks are also extended to the Speakers, who have graciously shared their time with us, to the Department of Highways for providing system support personnel and computer facilities, and to Mr. Macdonald and Dr. Cheng, as well as other staff members of the Treasury and Economics Department, for their help in making this seminar possible. I suggest that we give them a round of applause in appreciation.

APPENDIX

1. Agenda

2. Description of Tarela dictionary and General Background

3. Dictionary of retrieval information for test area

4. Examples of sample computer retrievals

5. DBS Brochure 'Geocoding-Facts by small Areas'

(The appendix consists of the documents which were distributed at the Seminar)

SEMINAR ON GEOCODING

PLACE:	St. Clair, Thames and Erie Room 2nd Floor, Macdonald Block, Queen's Park
DATE:	September 18, 1970
CHAIRMAN:	Mr. O.M. Schnick
9:00 a.m.	Welcome to the Seminar Mr. H.I. Macdonald Deputy Minister Department of Treasury and Economics
9:15 - 9:30	'General Background' - posing the problems of requirements for data on small area basis Dr. R.S. Thoman Director Regional Development Branch Department of Treasury and Economics
9:30 - 10:30	'Basic Concept and Geocoding by DBS' Mr. J.I. Weldon Coordinator, General Survey Systems Dominion Bureau of Statistics Ottawa
10:30 - 10:45	Coffee Break Humber Room
10:45 - 11:15	'Geocoding for Research' Professor R. McDaniel Department of Geography University of Western Ontario London, Ontario
11:15 - 12:00	'Distribution and Explanation of Computer Printou Mr. D. Weeks Programmer-Analyst Electronic Computing Branch Department of Highways
12:00 - 1:00	Lunch Humber Room, South
1:00 - 1:45	'Geocoding in the United States' Professor E.M. Horwood Professor of Civil Engineering University of Washington Seattle, Washington

1:45 - 2:15	'Rural and Single Property Geocoding' Mr. D.C. Symons Chief of Computer Services National Capital Commission Ottawa.
2:15 - 2:30	'Geocoding Activities in the Ontario Statistical Centre' Dr. K. Cheng Director Ontario Statistical Centre Department of Treasury and Economics
2:30 - 2:45	Coffee Break Humber Room
2:45 - 4:30	Panel Discussion - with audience participation Dr. K. Cheng (Chairman) Mr. J.I. Weldon Mr. A.E. Goodwin Professor R. McDaniel Professor E.M. Horwood Mr. D.C. Symons
4:30	Closing remarks by the Chairman Mr. O.M. Schnick Executive Director Economic and Statistical Services Division Department of Treasury and Economics

GEOCODING SEMINAR

September 18, 1970

General Background

This seminar is intended to bring together people interested in geocoding in Ontario. A description of a typical geocoding system was included with the invitation. This system, developed by the Dominion Bureau of Statistics, will be available for use in Ontario.

A team from the Ontario Government has been formed to investigate the possible use of the DBS Geocoding System by Ontario. The Dominion Bureau of Statistics will supply the programs as they are developed. The Ontario Statistical Centre, which is part of the Department of Treasury and Economics, will deal with potential users of the system, and will administer the use of the system. The Electronic Computing Branch of the Department of Highways of Ontario will implement the system on its computer for purposes of study. In the event that the DBS Geocoding System is adopted on a large scale, the Department of Highways could act as a service agency to operate and maintain the system.

The current status of the system is as follows: The Dominion Bureau of Statistics is nearing the completion of several years of research and development work on the Geocoding System. They have turned over to the Department of Highways about half of the system and this part is now operational on the DHO computer. The remainder of the system is expected to be released by DBS in the near future; possibly by the end of the year.

To-day's seminar will include a demonstration of information retrieval using part of the DBS system. This demonstration is intended to illustrate the simplicity of the retrieval language. The ability to retrieve data by user specified areas is not yet available. However, the retrieval language will remain the same when this feature is implemented in the near future. The language is introduced on the following pages. A more complete instruction manual on the use of the language is available from the Ontario Statistical Centre.

The participants in the seminar are invited to contact the Ontario Statistical Centre to discuss matters pertaining to Geocoding.

How to Prepare a Request

A request is formulated on a series of 80 column punch cards, using card columns 1-72. The format for these cards is discussed in 9.0.2.

A request consists of up to six parts. Each part may use more than one card if required. Each part begins with, and is distinguished by a keyword. The keywords are -

FILENAME:

AREANAME: (STATPAK only)

HEADING: (Optional, it may be omitted)

SELECTION CRITERIA: (optional)

CHARACTERISTICS:

TABULATE:

Each of these keywords has an acceptable abbreviation which may be used in lieu of spelling out the whole word as:

Column 1	F:	
	A:	(STATPAK only)
	H:	(Optional)
	S:	(Optional)
	C:	
	T:	

In all requests the keywords FILENAME: (or F:), CHARACTERISTICS: (or C:) and TABULATE: (or T:) must be present.

For STATPAK requests AREANAME: (or A:) must also be present.

SAMPLE REQUEST

FILENAME: HAMASSMT;

HEADING: THIS TABULATION SHOWS VARIOUS SUMS FOR OCCURRENCES OF AREA RANGES.

CHARACTERISTICS: ACRES UNDER 1.0, 1.01 to 10.0, OVER 10.0, TOTAL;

TABULATE: PRINT SUM (LAND), SUM (BLDG), SUM (TOTAL), SUM (GRFLR), PRINT COUNT;

OUTPUT

H: THIS TABULATION SHOWS VARIOUS SUMS FOR OCCURRENCES OF AREA RANGE

SUM(LAND) SUM(BLDG) SUM(TOTAL) SUM(GRFLR)	ACRES LE 1.0	ACRES 1.01-10.0	ACRES GT 10.0	ACRES TOTAL
	459,550	49,740	37,990	547,280
	1,616,590	26,710	0	1,643,300
	2,076,140	76,450	37,990	2,190,580
	267,599	3,000	0	270,599

COUNT	ACRES LE 1.0	ACRES 1.01-10.0	ACRES GT 10.0	ACRES TOTAL
	707	1	1	709

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL FORMAT	ORIGINAL CODE	WORD OR CODE	INTERPRETATION
County	CØUNTY	C2	25	WENTWØRTH	
Municipality	MUNIC	C2	18	HAMILTON	
Ward Number	WARD	C2	01-09		
Block Number	BLØCK	C3	001-113		
Book Number	BØØK	C2	00-99		
Base Number	BASE	C4	0000-9999		
Tenant Number	TENANT	C4	0000-9999		
Page Number	PAGE	C2	01-15		
			00		
			VO		
Polling Division	PØLDIV	C4	Blank		
NAME #1 Name	NAME1	C28	any /360 code		Assessed Person's Name
Marital Status	MS1	C2	Mb	MARRIED MAN	Married or Separated
			MW	MARRIED WØMAN	Married or Separated
			WR	WIDØWER	
			Wb	WIDØW	
			Bb	WINGLE MAN	Single or Divorced
			Sb	SINGLE WØMAN	Single or Divorced
			Øb		Meaning Unknown
			HW		Meaning Unknown
			bØ		Meaning Unknown
			bb		Meaning Unknown

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL FORMAT	ORIGINAL CODE	WORD OR CODE	INTERPRETATION
Alien	ALN1	C1	A	ALIEN	
			M		
			Blank	NØN_ALIEN	
Electoral Status	ES1	C2	Øb	OWNER	
			Lb	LESSEE	
			Tb	TENANT	
			ØC	ØCCUPIER	
			MF	MUNIC_FRANCHISE	
			FS	FARMERS_SON	
			FD	FARMERS_DAUGHT	
			SF	SIST_UNMAR_FARM	Sister of Unmarried Farmer
			EF	EXTEND_FRANCHISE	
			bb		Meaning Unknown
Religion	RELG1	C2	AN,AM,AS, BA,BH,BR, BU,CC,CH, CR,CS,DC, EV,GC,GØ bb,GT,JW, HE,LS,LU, ME,MØ,PA, PI,PN,PR, RC,SA,SD, SP,UC,UT, MS, LA,bO,ST		
Occupation	ØCUP1	C11	Alphabetic or blank		Occupation of person
Year of Birth	BRTH1	C2	00-99 or blank		Last two digits of year of birth

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL		WORD OR CODE	INTERPRETATION
		FORMAT	CODE		
Sex	SEX1	C1	M F Blank	MALE FEMALE	
Jury	JURY1	C1	J Blank	YES NØ	
School	SCHL1	C1	P S Blank B	PUBLIC SEPARATE	Public School support Separate School support Meaning Unknown Meaning Unknown
Acres	ACRES	P'9999V99T'	Numeric		Total acres of each property
Frontage	FRØNT	P'9999V99T	Numeric		Total frontage of each property
Depth	DEPTH	P'9999V99T	Numeric		Depth of each property
Res Pub	RESPUB	P'99999999T'	Numeric		Records the $amount of assesment which is liable for, Public School & Private school rate
Res Sep	RESSEP	P'99999999T'	Numeric		
Com Pub	CØMPUB	P'99999999T'	Numeric		Same, but for commercial
RCE (rate at which tax is paid)	RCE	C1	R	RESID_AND_FARM	Residential and farm rate Commercial and Industrial rate
	R-payment at residential rate		R	RESID_AND_FARM	

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL FORMAT	ORIGINAL CODE	WORD OR CODE	INTERPRETATION
RCE	C-payment at commercial rate		C	CØMM_AND_INDUST	Commercial and Industrial rate
	E-exempt		E	EXEMPT	Exempt from rates
			Blank		Meaning Unknown
Property Class	PRØPCL	C4	as per attached		To identify assessable units according to classifications required to complete analysis of assessment
School Support	PRØPSS	C1	P	PUBLIC	Public school support
			S	SEPARATE	Separate school support
			Blank		Meaning unknown
Corporation Code	CØRPCD	C1	C	CØRPØRATION	Assessable as a corporation
			Blank	NØT_CØRPØRATION	Not assessable as a corporation
ØCC Land	ØCCLND	P'99999999T'	Numeric		The $ amount of land assessment attributable to each unit
ØCC Bldg	ØCCBLD	P'99999999T'	Numeric		The $ amount of land assessment attributable to each building
ØCC Total	ØCCTØT	P'99999999T'	Numeric		Total
Business Percent	BUSPCT	C3	000-999 or MIN		
Business Amount	BUSAMT	P'99999999T'	Numeric		

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL		WORD OR CODE	INTERPRETATION
		FORMAT	CODE		
Number Employed	NUMEMP	C5	00000-99999		
Parking at Business	PARK	C4	0000-9999		
Cars	CARS	C2	00-99		
Trucks	TRUCKS	C2	00-99		
PI. Div.	PIDIV	C4	0000-9999		
Census Tract	CENTRK	C3	Numeric or blank		
Census Block	CENBLK	C3	Numeric or blank		
Economic Activity Classification	ECØN	C4	R000-R999 C000-C999 0000-9999 Blank,RØ71,		As per 1969 data collection code book Meaning Unknown Meaning Unknown
Ground Floor Area	GRFLR	C9	Numeric		Area of land building covers
Total Floor Area of all floors	TØTFLR	C9	Numeric		Area of all floors (includes apts,duplex, individual units).
Effective Year	YREFF	C3	000-999 or blank	uses only last 2 digits '70'	Year of construction
Number of Storeys	STØRIE	C3	000-999 00H-99H 00S-99S 00N		Number of storeys 'H' indicates ½ storey 'S' indicates a split No storeys

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL		WORD OR CODE	INTERPRETATION
		FORMAT	CODE		
RENTAL-UNFURNISHED U or F	UFU	C1	U F Blank	UNFURNISHED FURNISHED	
Year	YEARU	C2	00-99 or blank		Last two digits of Year
Actual	ACTU	C7	Numeric	$ amount per month of rent paid	
Estimated	ESTU	C7	Numeric	$ amount per month of rent estimated	
Light	LIGHTU	C1	X Blank	TENANT PAID LANDLØRD PAID	WHo Pays for lights
Heat	HEATU	C1	X or blank	As for Light	
Water	WATERU	C1	X or blank	As for Light	
Taxes	TAXESU	C1	X or blank	As for Light	
Linen	LINENU	C1	X or blank	As for Light	
Parking	PARKU	C1	X or blank	As for Light	
Other	OTHERU	C1	X or blank	As for Light	
Ratio	RATIØU	C4	Numeric	Ratio of rent to assessment	

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL FORMAT	ORIGINAL CODE	WORD OR CODE	INTERPRETATION
RENTAL-FURNISHED					
Same as Rental-Unfurnished. Last letter of name field is changed to 'F' i.e. Name for Year Furnished is YEARF					
Dwelling	DWLLG	C1	S Blank	SINGLE ØTHER	Single family house Other type of residential structure
Base	BASEBL	C3	X, Numeric, or blank		# of rooms in basement
Ground	GR	C3	X, Numeric, or blank		# of rooms on ground floor level
Second	SECD	C3	X, Numeric, or blank bb7		# of rooms on second floor level
Attic	ATTIC	C3	X, Numeric or blank		# of rooms in attic
Total	TØTBLD	C3	X, Numeric or blank		# of rooms
Bachelor	BACH	C3	X, Numeric or blank		# of bedrooms
One Bedroom	BR1	C3	X, Numeric or blank		
Two Bedrooms	BR2	C3	X, Numeric or blank		
Three Bedrooms	BR3	C3	X, Numeric or blank		

ASSESSMENT DICTIONARY

DESCRIPTION	NAME	ORIGINAL		WORD OR CODE	INTERPRETATION
		FORMAT	CODE		
Four Bedrooms	BR4	C3	X, Numeric or blank		
X-Co-ordinate	XCENT	C5	Numeric		
Y-Co-ordinate	YCENT	C6	Numeric		

PROPERTY CLASS

CODING SYSTEM

Code	Classification
BTWM	telephone wire mileage
BTGR	telephone gross receipts
TWM	telegraph wire mileage
TGR	telegraph gross receipts
TL	oil and gas transmission lines
RR	railway companies
C	commercial includes also: taxable institutional property; professional; gas land other than wells; gas wells; oil and gas distribution lines;
I	manufacturing and industrial
R	residential includes also: taxable property of churches, synagogues, convents, monasteries, etc. railway-residential assessment; conservation authority - taxable; taxable unfinished buildings; vacant residential (land plus improvement)
RBSU	residential basic shelter unit (multiple dwellings excluded)
RBSM	residential basic shelter multiple dwellings units
RL	residential land - unimproved farm includes also: woodland or forest; vacant farm (land plus improvement)
FBSU	farm basic shelter unit
FL	farmland - unimproved
W	wastelands
VC	all properties of a commercial and professional nature not subject to business assessment includes also: vacant vacation resort
VCL	commercial land - unimproved
VI	all properties of a manufacturing and industrial nature not subject to business assessment
VIL	industrial land - unimproved
FGL	Federal Government - liable for grants
FGAL	Federal Government Agency - liable for grants

SAMPLE REQUEST

```
F:     HAMASSMT;

H:     NUMBER OF TENANTS AND OWNERS BY TYPE OF TENURE
       WITH 2 BEDROOMS;

C:     ES1 'T', 'O';
       BR2 NE  ' ';
       TENURE '0'-'4', '5'-'8', '9';

T:     COUNT;
```

OUTPUT

H: NUMBER OF TENANTS AND OWNERS BY TYPE OF TENURE WITH 2 BEDROOMS;

COUNT

		TENURE '0'-'4'	TENURE '5'-'8'	TENURE '9'
ES1 BR2	'T' NE ' '	27	0	0
ES1 BR2	'O' NE ' '	29	0	0

SAMPLE REQUEST

```
F:      HAMASSMT;

H:      COUNT NON-ALIEN, MARRIED MALE OWNERS OVER 21
        WHO ARE ELIGIBLE FOR JURY DUTY (NAME1 ONLY);

C:      MS1 'M';
        ALN1 ' ';
        BRTH1   LT '20', '20'-'30', '31'-'40', '41'-'48';
        JURY1 'J';

T:      COUNT;
```

OUTPUT

```
H:    COUNT NON-ALIEN, MARRIED MALE OWNERS OVER 21 WHO ARE
      ELIGIBLE FOR JURY DUTY (NAME1 ONLY);

COUNT
                                 |   JURY1      |
                                 |   'J'        |
                                 |              |
-------------------------------------------------
                                 |              |
                                 |              |
                                 |              |
MS1          'M '                |              |
 ALN1         ' '                |              |
  BRTH1        LT '20'           |            73|
  BRTH1        '20'-'30'         |            53|
  BRTH1        '31'-'40'         |            47|
  BRTH1        '41'-'48'         |            37|
                                 |              |
                                 |              |
```

SAMPLE REQUEST

```
F:      HAMASSMT;

H:      LAND USES AND DOLLAR LAND ASSESSMENT IN MUNIC. 18,
        VS. POPULATION;

C:      MUNIC  '18';
        POP '0'-'500',  '501'-'1000',  '1000'-'5000',
        GT '5000';
        LNDUSE 'R', 'A', 'C', 'I';

T:      PRINT COUNT, AVERAGE (ACRES), AVERAGE (LAND),
        SUM (ACRES), SUM (LAND), PRINT COUNT;
```

OUTPUT

H: LAND USES AND DOLLAR LAND ASSESSMENT IN MUNIC. 18, VS. POPULATION

COUNT
AVERAGE (ACRES)
AVERAGE (LAND)
SUM (ACRES)
SUM (LAND)

		LNDUSE 'R'	LNDUSE 'A'	LNDUSE 'C'	LNDUSE 'I'
MUNIC	'18'				
POP	'0'-'500'	262	5	11	5
		0	0	0	3
		660	11,076	16,857	19,104
		0	0	0	16
		173,060	55,380	185,430	95,520
POP	'501'-'1000'	0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0
POP	'1001'-'5000'	0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0
POP	GT '5000'	0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0
		0	0	0	0

H: FRI., SEPT. 18, 1970,
REQUEST #1,
NUMBER OF SINGLE AND MARRIED MEN BY BIRTH AND ELECTORAL STATUS (OWNER, TENANT, MUNIC. FRANCHISE);

COUNT

		BRTH1 '40'-'45'
ES1	'O'	
MS1	'B'	0
MS1	'M'	0
ES1	'T '	
MS1	'B '	13
MS1	'M '	35
ES1	'MF'	
MS1	'R '	0
MS1	'M '	0

FRI., SEPT. 18, 1970,
REQUEST #2,
NUMBER OF OWNERS AND TENANTS SUPPORTING PUBLIC AND SEPARATE SCHOOLS;

COUNT

		SCHL1 'P'	SCHL1 'S'
ES1	'O '	0	0
ES1	'T '	293	62

H: FRI., SEPT. 18, 1970,
REQUEST #3,
NUMBER OF OCCURRENCES OF CORPORATIONS WITH SHOWN TRUCK NO. RANGE;

S: CORPCD 'C';

COUNT

	TRUCKS '00'	TRUCKS '01'-'05'	TRUCKS GF '06'	TRUCKS TOTAL

H: FRI., SEPT. 18, 1970,
REQUEST #3,
NUMBER OF OCCURRENCES OF CORPORATIONS WITH SHOWN TRUCK NO. RANGE;

S: CORPCD 'C';

COUNT

	TRUCKS '00'	TRUCKS '01'-'05	TRUCKS GE '06'	TRUCKS TOTAL
	28	4	3	35

geocoding FACTS BY SMALL AREAS

A new method of assembling census and other DBS data, providing more flexible and extensive availability of information by user-specified areas.

Bulletin #1 February, 1969

The Dominion Bureau of Statistics has under development a computerized system for providing census data for 1971 on a user-specified basis in large urban areas — and in some other areas.

It is envisaged that the system will eventually make available any combination of census data for virtually any area that the user might specify (within minimal limitations). The main objective is to provide tabulations relatively quickly and inexpensively by automatic selection and aggregations of a series of building blocks that would make up the user-specified area. In the larger urban areas these would be city block faces; enumeration areas will be used elsewhere. The service would be made possible by the automatic precoding of all census addresses.

Eventually, divers socio-economic statistics from other surveys could become available on a similar basis, with cross-tabulations in a variety of combinations.

This is the first in a series of bulletins to inform users of census data on plans and progress in the development of a DBS Geographically Referenced Data Storage and Retrieval System (GRDSR). This system, designed to meet the growing information needs of administrators, planners and researchers in the social, economic, business and other fields, should be particularly valuable to planners, developers and users of municipal management information systems. It could also offer important benefits for many other types of users.

GRDSR places the emphasis on making information available in larger urban areas by user-specified areas, as opposed to standard areas such as enumeration areas, census tracts, and municipalities (for which, however, census data will continue to be provided).

The system consists of a set of data processing operations and the storage and retrieval of corresponding data on randomly accessible data storage devices. It provides flexibility for the retrieval and tabulation of any combination of census data and for cross-referencing of different data files by any user-specified area (provided always that the confidentiality requirements of the Statistics Act are safeguarded).

GRDSR, which is the outcome of two years of research, has been designed specifically for larger urban areas for the 1971 census. Less extensive but similar capabilities are planned for the rest of the country. Although designed initially for manipulating data derived from population censuses, the system may also be extended to manufacturing, retail and agricultural census data.

It is being developed in response to increasing demands on DBS – which the Bureau cannot now economically service – for tabulations of statistics arranged by other than standard geographical areas (e.g. census tracts).

DBS, as Canada's central statistical agency, attaches great importance to achieving the capability to meet such demands and making intra-urban information available on a uniform, national basis.

Published under the authority of the Minister of Industry, Trade and Commerce

DOMINION BUREAU OF STATISTICS, OTTAWA 3, ONTARIO

Conceptual Aspects

GRDSR is based on the fact that most DBS surveys have common reference points – the addresses of respondents, which can be given geographical coordinates.

On this basis, once a survey (census, for example) is taken, the data obtained from each respondent, with his address, can be converted to a machine readable form. Then the appropriate geographical coordinate, as referenced in the Universal Transverse Mercator System, is linked to the address and automatically replaces it.

A basic requirement is an address conversion file. This lists all block faces (generally one side of a street between neighbouring intersections) by street names, by block face terminal addresses, and by corresponding centroid coordinates. An essential working machine readable file, it must be kept constantly up to date as to changes in addresses, changes in street names and all other pertinent data.

How Users May Define Areas

Using the block faces as building blocks, the urban user can define his own specific study area simply by outlining the block faces within the desired area. This may be done, and preferably should be done, on a computer-printed map which the Bureau proposes to supply.

Areas may be enclosed by streets, or by other well-defined boundaries, may cut across boundary lines of census tracts or enumeration areas (in urban applications) but may not cut across block faces. Thus the user has very considerable flexibility in areal delineation and almost unlimited practical possibilities are opened up for users whose interests are essentially small area in nature. Typical of areas that could be studied under the GRSDR system are school districts, town planning districts, traffic zones, product testing and marketing zones.

It must be noted, however, that the constraint of Statistics Act confidentiality requirements – which prohibits disclosure of information on individuals or individual bodies – remains. The user should not, therefore, expect to receive data for individual block faces or even city blocks.

Benefits of the system can, however, far outweigh this constraint.

Among these benefits is that the technique might be equally usable for locally available computerized municipal data. Arrangements may be possible for local agencies to obtain the computer programs used by the GRDSR system to be locally operated on other than DBS data.

Storing Data

Once geocoded, census data for individual records are stored as strings – each string recording the information for one data characteristic for the population reported.

Information in each string will be arranged as to:

- Individuals within households.
- Households within block faces.
- Block faces within the urban geocoded area.

There are as many data strings as there are data characteristics recorded. While the design of the data strings assures maximum efficiency in retrieval and cross-tabulating of data, the required data strings and their portions corresponding to the designated retrieval area are accessed through the block face centroids.

Retrieving Data

By this means of storing data it is expected that retrieval will be a relatively simple operation.

The initial step will be for the user to specify exact data characteristics and the precise variables for these characteristics (as in age, sex, income, ethnic origin) and the boundaries of the requested area.

Computer processing will then, as a first step, select all the block face centroids which lie within the area. From this point, a generalized program will retrieve and tabulate requested data fields bearing the selected block face identifications. No programming work will be required on the part of the user, nor any knowledge of computer programming.

The Scope — and Limitations

Geocoding of urban areas requires a large initial supply of street input information such as accurate street maps and up-to-date address ranges – and this information must be kept constantly updated.

Since this information must be coded for computer processing, there are obvious limits on the number of urban areas that can be geocoded for the 1971 census. Present objectives call for geocoding those areas that had a population in the city proper in 1966 of at least 100,000 – providing also that there are local agencies in these areas that are prepared to supply and periodically update the required street input information.

Plans For Other Areas

An alternative form of geocoding, based for the most part on assigning geographic coordinates to enumeration areas, is also planned for 1971 in all areas not otherwise geocoded. This will cover many areas that are obviously urban in character and which, in time, will be refined to a block face level.

Local Participation

Municipalities generally appear willing to work jointly with DBS toward attainment of the common objective – the availability of more flexible data – and the degree of

their willingness to assist in supplying street input information is a determinant of achieving geocoding in their areas.

Their participation is a logical contribution. Local agencies are most familiar with their areas and have an obvious self-interest in establishing an automated, updatable, nationally compatible urban data system that can be queried for short and long range decision making.

The first contribution sought by DBS is, of course, source documents (basically, maps and address ranges by block faces), checking of discrepant information and a continuing supply of update information and, perhaps, coding of street pattern information – all preferably through one designated agency for the urban area concerned.

In return for such participation, DBS would be in the position to provide tabulations from the 1971 census by user specified areas in the locality concerned. DBS also expects to offer the local agency access to the computer programs necessary to geocode their own data and to retrieve such data for any query area.

Such programs would be designed to operate on the type of medium-scale computer the agency might have or would be available in a nearby service bureau. These programs, typically, would enable the local agency or its clients to geocode, store and retrieve locally generated data covering such areas as assessment, planning, traffic, land utilization, zoning, education, health and welfare.

Tabulations from locally generated data could be supplemented with census data on an aggregate basis.

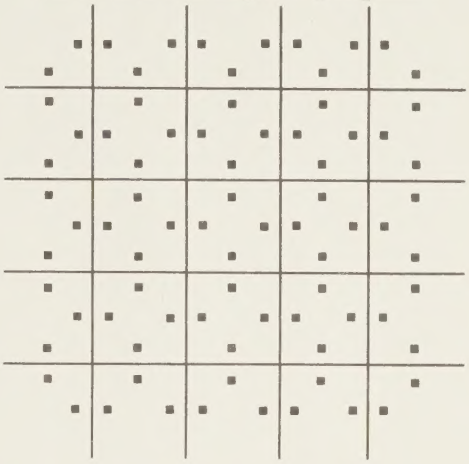
The Nature of the Need

The nature of the need for such data services was underlined in the 1966 census which showed that nearly one half of Canada's population at that time – some 9.7 million people – were then living in 19 metropolitan areas. These needs do not abate. The Economic Council of Canada has estimated that well over 80 per cent of the 25 million population it forecasts for Canada in 1980 will live in urban areas – and that about 40 per cent of these urban dwellers will live in the Montreal, Toronto, Vancouver, Winnipeg, Calgary, Edmonton and Ottawa regions alone.

The authorities responsible for the development of metropolitan areas are not unaware of their own need for gathering and computerizing data for planning purposes – and a multiplicity of computerized urban information systems could easily develop in the absence of close cooperation between the various levels of governments. Several cities may already have independent programs under way. These systems may not be compatible each with the other, however, thus creating problems in the effective exchange and utilization of information.

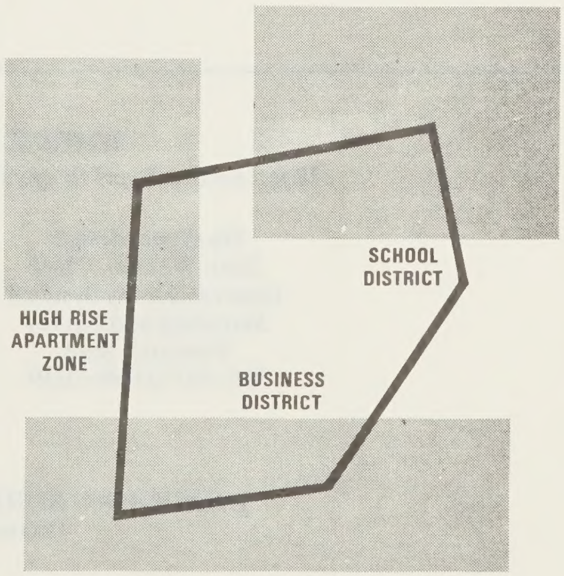
Typical Urban Retrieval Areas
(Dots represent block face centroids)

Where streets arranged in grid pattern



Where streets follow irregular pattern

How Typical Retrieval Areas May Cross Census Tract Boundaries
(Heavy line represents a theoretical urban census tract)



WHERE TO INQUIRE

If you need answers to specific questions on GRDSR, contact:

On system design:
John Weldon, Chief
General Survey Systems
Sampling and Survey
Research Staff
Tel: (613) 996-1039

On potential census applications:
W. D. Porter, Director
Census Division
Tel: (613) 994-9454

DOMINION BUREAU OF STATISTICS
Ottawa 3, Ontario